DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 15329683 Document Date: 01/27/2017

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

10x Genomics EX1004 (Part 1)

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 15329683 Document Date: 01/27/2017

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

PTO-1390 (06-13)
Approved for use through 6/30/2013. OMB 0651-0021
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A SUBMISSION UNDER 35 U.S.C. 371

Attorney Docket No.	
H0498.70509US02	
U.S. Application No. (if known, see 37 CFR 1.5)	

International Application No.	International Filing Date	Priority Date Claimed
PCT/US2015/042556	29 July 2015	30 July 2014

Title of Invention: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

First Named Inventor: Xiaowei Zhuang

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information.

1. ☐ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). NOTE: The express request under 35 U.S.C. 371(f) will not be effective unless the requirements under 35 U.S.C. 371(c)(1), (2), and (4) for payment of the basic national fee, copy of the International Application and English translation thereof (if required), and the oath or declaration of the inventor(s) have been received.

2. ☐ A copy of the International Application (35 U.S.C. 371(c)(2)) is attached hereto (not required if the International Application was previously communicated by the International Bureau or was filed in the United States Receiving Office (RO/US)).

3. An English language translation of the International Application (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto.
   b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).

4. An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4))
   a. ☐ is attached.
   b. ☐ was previously filed in the international phase under PCT Rule 4.17(iv).

Items 5 to 8 below concern amendments made in the international phase.

PCT Article 19 and 34 amendments

5. ☐ Amendments to the claims under PCT Article 19 are attached (not required if communicated by the International Bureau) (35 U.S.C. 371(c)(3)).

6. ☐ English translation of the PCT Article 19 amendment is attached (35 U.S.C. 371(c)(3)).

7. ☐ English translation of annexes (Article 19 and/or 34 amendments only) of the International Preliminary Examination Report is attached (35 U.S.C. 371(c)(5)).

Cancellation of amendments made in the international phase

8a. ☐ Do not enter the amendment made in the international phase under PCT Article 19.

8b. ☐ Do not enter the amendment made in the international phase under PCT Article 34.

NOTE: A proper amendment made in English under Article 19 or 34 will be entered in the U.S. national phase application absent a clear instruction from applicant not to enter the amendment(s).

The following items 9 to 17 concern a document(s) or information included.

9. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

10. ☒ A preliminary amendment.

11. ☒ An Application Data Sheet under 37 CFR 1.76.

12. ☐ A substitute specification. NOTE: A substitute specification cannot include claims. See 37 CFR 1.125(b).

13. ☐ A power of attorney and/or change of address letter.

14. ☒ A computer-readable form of the sequence listing in accordance with PCT Rule 13*ter*.3 and 37 CFR 1.821-1.825.

15. ☐ Assignment papers *(cover sheet and document(s))*. Name of Assignee: _____

16. ☐ 37 CFR 3.73(c) Statement *(when there is an Assignee)*.

5220196.1

PTO-1390 (06-13)
Approved for use through 6/30/2013. OMB 0651-0021
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

U.S. APPLN. No. (if known – see 37 CFR 1.5)	INTERNATIONAL APPLICATION No.	ATTORNEY DOCKET No.
	PCT/US2015/042556	H0498.70509US02

17. [X] Other items or information: PCT Published Application, International Search Report, Invitation to Pay Additional Fees, Form PCT/IB/304

The following fees have been submitted.	CALCULATIONS	PTO USE ONLY
18. [X] Basic national fee (37 CFR 1.492(a)) **$280**	$ 280.00	
19. [X] Examination fee (37 CFR 1.492(c)) If the written opinion prepared by ISA/US or the international preliminary examination report prepared by IPEA/US indicates all claims satisfy provisions of PCT Article 33(1)-(4) ..... **$0** All other situations **$720**	$ 720.00	
20. [X] Search fee (37 CFR 1.492(b)) If the written opinion prepared by ISA/US or the international preliminary examination report prepared by IPEA/US indicates all claims satisfy provisions of PCT Article 33(1)-(4) ..... **$0** Search fee (37 CFR 1.445(a)(2)) has been paid on the international application to the USPTO as an International Searching Authority ..... **$120** International Search Report prepared by an ISA other than the US and provided to the Office or previously communicated to the US by the IB ..... **$480** All other situations **$600**	$ 120.00	
TOTAL OF 18, 19, and 20 =	$ 1,120.00	
[X] Additional fee for specification and drawings filed in paper over 100 sheets (excluding sequence listing in compliance with 37 CFR 1.821(c) or (e) in an electronic medium or computer program listing in an electronic medium) (37 CFR 1.492(j)). Fee for each additional 50 sheets of paper or fraction thereof **$400**		

Total Sheets	Extra Sheets	Number of each addition 50 or fraction thereof (round **up** to a whole number)	RATE	
159 - 100 =	59 / 50 =	2	x **$400**	$ 800.00

Surcharge of **$140.00** for furnishing any of the search fee, examination fee, or the oath or declaration after the date of commencement of the national stage (37 CFR 1.492(h)).	$

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	
Total claims	- 20 =	0	x **$80**	$
Independent claims	- 3 =	0	x **$420**	$
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ **$780**	$

Processing fee of **$140.00** for furnishing the English translation later than 30 months from the earliest claimed priority date (37 CFR 1.492(i)). +	$
TOTAL OF ABOVE CALCULATIONS =	$ 1,920.00
[X] **Applicant asserts small entity status.** See 37 CFR 1.27. Fees above are reduced by ½.	- 960.00
[ ] **Applicant certifies micro entity status.** See 37 CFR 1.29. Fees above are reduced by ¾. Applicant must attach form PTO/SB/15A or B or equivalent.	
TOTAL NATIONAL FEE =	$ 960.00
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property. +	$
TOTAL FEES ENCLOSED =	$ 960.00
Amount to be refunded:	$
Amount to be charged:	$

a. [ ] A check in the amount of $ _____ to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. 23/2825 in the amount of $ _____ to cover the above fees.

c. [X] The Director is hereby authorized to charge additional fees which may be required, or credit any overpayment, to Deposit Account No. 23/2825 as follows:

i. [ ] any required fee.

ii. [X] any required fee except for excess claims fees required under 37 CFR 1.492(d) and (e) and multiple dependent claim fee required under 37 CFR 1.492(f).

d. [X] Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038. The PTO-2038 should only be mailed or faxed to the USPTO. However, when paying the basic national fee, the PTO-2038 may NOT be faxed to the USPTO.

ADVISORY: If filing by EFS-Web, do **NOT** attach the PTO-2038 form as a PDF along with your EFS-Web submission. Please be advised that this is **not** recommended and by doing so your **credit card information may be displayed via PAIR**. To protect your information, it is recommended to pay fees online by using the electronic payment method.

NOTE: Where an appropriate time limit under 37 CFR 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the International Application to pending status.

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File)Transition Applications

[ ] This application (1) claims priority to or the benefit of an application filed before March 16, 2013, and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.

NOTE 1: By providing this statement under 37 CFR 1.55 or 1.78, **this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA**.

NOTE 2: A U.S. national stage application may not claim priority to the international application of which it is the national phase. The filing date of a U.S. national stage application is the international filing date. See 35 U.S.C. 363.

Correspondence Address

[X] The address associated with Customer Number: 86110 **OR** [ ] Correspondence address below

Name					
Address					
City		State		Zip Code	
Country		Telephone			
Email					

Signature	/Tani Chen/	Date	January 27, 2017
Name (Print/Type)	Tani Chen, Sc.D.	Registration No. (Attorney/Agent)	52,728

Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Xiaowei
Family Name::	Zhuang
City of Residence::	Lexington
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	5 Luongo Farm Lane
City of mailing address::	Lexington
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02421

Inventor Number::	2
Given Name::	Kok-Hao
Family Name::	Chen
City of Residence::	Singapore
Country of Residence::	Republic of Singapore
Street of mailing address::	37 Jalan Lepas
City of mailing address::	Singapore
Country of mailing address::	Republic of Singapore
Postal or Zip Code of mailing address::	537961

Inventor Number::	3
Given Name::	Alistair
Family Name::	Boettiger
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	1550 Cambridge Street
	Apt. #2
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02139

Inventor Number::	4
Given Name::	Jeffrey
Middle Name::	R.
Family Name::	Moffitt
City of Residence::	Somerville
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	175 Beacon Street
	Apt. #108

City of mailing address::	Somerville
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02143

Inventor Number::	5
Given Name::	Siyuan
Family Name::	Wang
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	28 Fernald Drive
	Apt. #31
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Correspondence Information

Correspondence Customer Number::	86110

Application Information

Application Type::	Regular
Subject Matter::	Utility

CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Attorney Docket Number::	H0498.70509US02
Request for Early Publication?::	No
Request for Non-Publication?::	No
Small Entity?::	Yes
Petition included?::	No
Licensed US Govt. Agency::	National Institutes of Health
Contract or Grant Numbers::	GM096450
Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2::	No

Representative Information

Representative Customer Number::	86110

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::	Prior Appl Status::
This Application	National Stage of	PCT/US2015/042556	07/29/15	
PCT/US2015/042556	An application claiming the benefit under 35 USC 119(e)	62/142653	04/03/15	
PCT/US2015/042556	An application claiming the benefit under 35 USC 119(e)	62/050636	09/15/14	
PCT/US2015/042556	An application claiming the benefit under 35 USC 119(e)	62/031062	07/30/14	

Foreign Priority Information

Applicant Information

Applicant Number::	1
Applicant Type::	Assignee
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Assignee Information Including Non-Applicant Assignee Information

Assignee Number::	1
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Country of mailing address::	US
Postal or Zip Code of mailing address::	02138

Authorization or Opt-Out of Authorization to Permit Access

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

[X] A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

[X] B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

Signature:

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Tani Chen/	Date (YYYY-MM-DD)	2017-01-27
Name	Tani Chen, Sc.D.	Registration Number	52,728

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	National Stage of PCT/US2015/042556
Confirmation No.:	Not Yet Assigned
Filed:	Concurrently Herewith
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	Not Yet Assigned
Art Unit:	Not Yet Assigned

PRELIMINARY AMENDMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

Prior to examination on the merits, please amend the above-identified U.S. patent application as follows:

Amendments to the Specification begin on page 2 of this paper.

Remarks begin on page 3 of this paper.

AMENDMENTS TO THE SPECIFICATION

Please amend the "Related Applications" section on page 1, lines 3-10 as follows. Additions are indicated by underlining and deletions are indicated by strikeouts and/or double bracketing.

This application is a national stage filing of International Patent Application Serial No. PCT/US2015/042556, which claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, et al.; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, et al.; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, et al. Each of the above is incorporated herein by reference.

REMARKS

This is a Preliminary Amendment in which the priority claim has been updated to reflect national entry of this application.

Dated: January 27, 2017

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
Timothy J. Oyer, Ph.D.
Registration No.: 36,628
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 February 2016 (04.02.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/018960 A1

(51) International Patent Classification:
C40B 30/04 (2006.01) *C12Q 1/68* (2006.01)
C07H 21/02 (2006.01)

(21) International Application Number:
PCT/US2015/042556

(22) International Filing Date:
29 July 2015 (29.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/031,062 30 July 2014 (30.07.2014) US
62/050,636 15 September 2014 (15.09.2014) US
62/142,653 3 April 2015 (03.04.2015) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **ZHUANG, Xiaowei**; 5 Luongo Farm Lane, Lexington, MA 02421 (US). **CHEN, Kok-Hao**; 37 Jalan Lepan, Singapore 537961 (SG). **BOETTIGER, Alistair**; 1550 Cambridge Street, Apt. #2, Cambridge, MA 02139 (US). **MOFFITT, Jeffrey, R.**; 175 Beacon Street, Apt. #108, Somerville, MA 02143 (US). **WANG, Siyuan**; 28 Fernald Drive, Apt. #31, Cambridge, MA 02138 (US).

(74) Agent: **CHEN, Tani**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

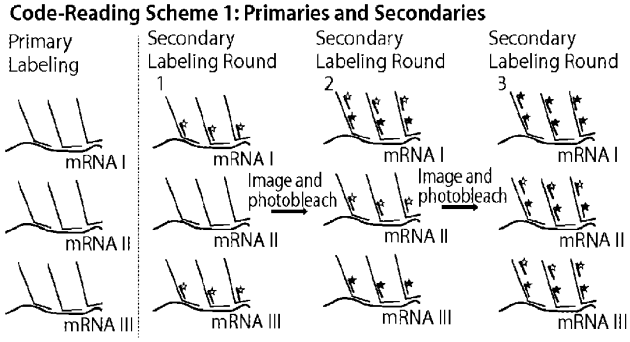


Fig. 1B

(57) Abstract: The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

WO 2016/018960 A1

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first

secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Fig. 12 shows schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16B show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention; and

Figs. 19A-9D illustrate decoding and error assessment, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:
GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA
TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG
GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;

SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;

SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;

SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and

SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the

cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences.

For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize

a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was

shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that

allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%,

at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell

may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40,

at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For

instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at

least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments,

the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting

code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this

number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or

more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about 500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about

0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification

and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic

acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities

are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another

set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light

having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be

connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate,

Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso

Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate

control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled "Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate,

polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those

sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three

Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the

bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs

computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps (meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 °C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected

while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 ^{o}C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 ^{o}C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final

1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics

setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0]. Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not

achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as N increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper

symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse

transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in *N* rounds of imaging. Each RNA species is encoded with a *N*-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (*N*) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of *N* different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the *N* readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. *N* subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb *N*). The bound probes were inactivated by photobleaching between successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA

species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction. To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes

containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Fig. 20). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control

words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Fig. 20) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range

additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of $1x10^{-53}$. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads (FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of $3x10^{-39}$.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4

code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e., a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification

control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Fig. 20 shows two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated

or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1, PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by

increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient $r = 0.76$; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data ($r = 0.65$; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of $3x10^{-133}$. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of $3x10^{-33}$. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of $9x10^{-30}$. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene

experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the

140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACCG (SEQ ID NO: 20)	TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC (SEQ ID NO: 23)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of

homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 °C; there is no cross hybridization targets with melting temperatures greater than 72 °C; there is no predicted internal secondary structures with melting temperatures greater than 76 °C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using

BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 °C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 °C for 45 minutes, and the reverse transcriptase was inactivated at 85 °C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 °C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 °C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to

confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5 minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged

immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope

backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold

required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all

subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 ^{o}C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 ^{o}C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible code words, the calling rate, and the misidentification rate on N. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong

species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible N-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00...0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with m '1's and N–m '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \tag{1}$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \tag{2}$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with m '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \tag{3}$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the

fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m} + m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m} + (N-m)p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the m '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the N-m '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of m. The weight for each term was determined from the number of words that contain m '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with m '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^i p_0^{4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)} + \\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^i p_0^{3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)} + \\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^i p_0^{5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)} +
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have m '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did

not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}&\binom{4}{2}\binom{N-4}{2}p_1^2p_0^2(1-p_1)^2(1-p_0)^{N-5}+\\&\binom{4}{1}\binom{N-4}{2}p_1^1p_0^2(1-p_1)^3(1-p_0)^{N-6}+\binom{4}{2}\binom{N-4}{1}p_1^2p_0^1(1-p_1)^2(1-p_0)^{N-5}\\&\binom{4}{2}\binom{N-4}{3}p_1^2p_0^3(1-p_1)^2(1-p_0)^{N-7}+\binom{4}{3}\binom{N-4}{2}p_1^3p_0^2(1-p_1)^1(1-p_0)^{N-6}\end{aligned} \quad (6)$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be A, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \quad (7)$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in

order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and

optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

1. A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2. The method of claim 1, comprising exposing the sample to at least 5 different nucleic acid probes.

3. The method of any one of claims 1 or 2, comprising exposing the sample to at least 10 different nucleic acid probes.

4. The method of any one of claims 1-3, comprising exposing the sample to at least 100 different nucleic acid probes.

5. The method of any one of claims 1-4, comprising exposing the sample to a plurality of nucleic acid probes simultaneously.

6. The method of any one of claims 1-5, comprising exposing the sample to a plurality of nucleic acid probes sequentially.

7. The method of claims 1-6, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8. The method of claim 7, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of RNA species in the sample.

9. The method of claim 7 or 8, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of DNA sequences in the sample.

10. The method of any one of claims 1-9, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11. The method of claim 10, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

12. The method of any one of claims 10 or 11, wherein the target sequence is substantially complementary to a nucleic acid sequence encoding a protein.

13. The method of any one of claims 10-12, wherein the target sequence is substantially complementary to messenger RNA (mRNA).

14. The method of any one of claims 10-13, wherein the plurality of nucleic acid probes comprises at least 8 possible read sequences.

15. The method of any one of claims 10-14, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

16. The method of any one of claims 10-15, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

17. The method of any one of claims 10-16, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

18. The method of any one of claims 10-17, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19. The method of any one of claims 10-18, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

20. The method of any one of claims 10-19, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

21. The method of any one of claims 10-20, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

22. The method of any one of claims 10-21, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

23. The method of any one of claims 10-22, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

24. The method of any one of claims 10-23, wherein at least some of the plurality of nucleic acid probes comprise no more than 3 read sequences.

25. The method of any one of claims 10-24, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

26. The method of claims 1-25, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27. The method of claim 26, wherein the first signaling entity is fluorescent.

28. The method of any one of claims 26 or 27, wherein the first signaling entity is a protein.

29. The method of any one of claims 26-28, wherein the first signaling entity is a dye.

30. The method of any one of claims 26-29, wherein the first signaling entity is a nanoparticle.

31. The method of any one of claims 26-30, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32. The method of claim 31, wherein the first signaling entity and the second signaling entity are identical.

33. The method of claim 31, wherein the first signaling entity and the second signaling entity are not identical.

34. The method of any one of claims 31-33, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35. The method of claim 34, comprising inactivating the first signaling entity by photobleaching at least some of the first signaling entity.

36. The method of any one of claims 34 or 35, comprising inactivating the first signaling entity by chemically bleaching at least some of the first signaling entity.

37. The method of any one of claims 34-36, comprising inactivating the first signaling entity by exposing the first signaling entity to a reactant able to alter the structure of the signaling entity.

38. The method of any one of claims 34-37, comprising inactivating the first signaling entity by removing at least some of the first signaling entity.

39. The method of any one of claims 34-38, comprising inactivating the first signaling entity by dissociating the first signaling entity from the first secondary probe.

40. The method of any one of claims 34-39, comprising inactivating the first signaling entity by dissociating the first secondary probe that contains the first signaling entity from the sample.

41. The method of any one of the claims 34-40, comprising inactivating the first signaling entity by chemically cleaving it from the first secondary probe.

42. The method of any one of the claims 34-41, comprising inactivating the first signaling entity by enzymatically cleaving it from the first secondary probe.

43. The method of any one of claims 34-42, comprising inactivating the first signaling entity by exposing the signaling entity or the first secondary probe to an enzyme.

44. The method of any one of claims 1-43, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

45. The method of any one of claims 1-44, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

46. The method of any one of claims 44 or 45, comprising determining the centroid using a maximum likelihood algorithm.

47. The method of any one of claims 44-46, comprising determining the centroid using a least squares algorithm.

48. The method of any one of claims 44-47, comprising determining the centroid using a Bayesian algorithm.

49. The method of any one of claims 44-48, comprising determining the centroid using a compressed sensing algorithm.

50. The method of any one of claims 1-49, further comprising determining a confidence level for the identified nucleic acid target.

51. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having one or more one-bit errors to the codeword.

52. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having exactly one one-bit error to the codeword.

53. The method of claims 1-52, wherein at least some of the plurality of nucleic acid probes comprise DNA.

54. The method of any one of claims 1-53, wherein at least some of the plurality of nucleic acid probes comprise RNA.

55. The method of any one of claims 1-54, wherein at least some of the plurality of nucleic acid probes comprise PNA.

56. The method of any one of claims 1-55, wherein at least some of the plurality of nucleic acid probes comprise LNA.

57. The method of any one of claims 1-56, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

58. The method of any one of claims 1-57, wherein at least some of the plurality of nucleic acids probes are configured to bind to a nucleic acid within the sample.

59. The method of any one of claims 1-58, wherein at least some of the binding of the nucleic acid probes to a target within the sample is specific binding.

60. The method of any one of claims 1-59, wherein at least some of the binding of the nucleic acid probes to a target within the sample is via Watson-Crick base pairing.

61. The method of any one of claims 1-60, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

62. The method of any one of claims 1-61, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

63. The method of any one of claims 1-62, wherein at least some of the plurality of nucleic acid probes are configured to bind to mRNA.

64. The method of any one of claims 1-63, wherein at least some of the plurality of nucleic acid probes are configured to bind to transfer RNA (tRNA).

65. The method of any one of claims 1-64, wherein at least some of the plurality of nucleic acid probes are configured to bind to ribosomal RNA (rRNA).

66. The method of any one of claims 1-65, wherein at least some of the plurality of nucleic acid probes are configured to bind to lncRNA.

67. The method of any one of claims 1-66, wherein at least some of the plurality of nucleic acid probes are configured to bind to snoRNA.

68. The method of any one of claims 1-67, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

69. The method of any one of claims 1-68, wherein the plurality of nucleic acids probes is configured to bind to DNA.

70. The method of any one of claims 1-69, wherein at least some of the plurality of nucleic acid probes are configured to bind to genomic DNA.

71. The method of any one of claims 1-70, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

72. The method of any one of claims 1-71, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm.

73. The method of any one of claims 1-72, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 80 nm.

74. The method of any one of claims 1-73, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 50 nm.

75. The method of any one of claims 1-74, wherein the sample comprises a cell.

76. The method of claim 75, wherein the cell is a human cell.

77. The method of any one of claims 76 or 77, wherein the cell is fixed.

78. The method of any one of claims 1-77, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

79. The method of any one of claims 1-78, comprising determining binding of the nucleic acid probes using an optical imaging technique.

80. The method of any one of claims 1-79, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

81. The method of any one of claims 1-80, comprising determining binding of the nucleic acid probes using a multi-color fluorescence imaging technique.

82. The method of any one of claims 1-81, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

83. The method of claim 82, comprising determining binding of the nucleic acid probes using stochastic optical reconstruction microscopy (STORM).

84. The method of any one of claims 82 or 83, comprising determining binding of the nucleic acid probes using photoactivated localization microscopy (PALM) or fluorescence photoactivation localization microscopy (FPALM).

85. The method of any one of claims 82-84, comprising determining binding of the nucleic acid probes using stimulated emission depletion microscopy (STED).

86. The method of any one of claims 82-85, comprising determining binding of the nucleic acid probes using structured illumination microscopy (SIM).

87. The method of any one of claims 82-86, comprising determining binding of the nucleic acid probes using reversible saturated optically linear fluorescence transition (RESOLFT) microscopy.

88. A method, comprising:

exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and

for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

89. The method of claim 88, further comprising:

creating codewords based on determination of the read sequences within the sample.

90. The method of claim 89, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

91. A method, comprising:

exposing a sample to a plurality of primary nucleic acid probes;

exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample;

creating codewords based on fluorescence of the secondary nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

92. A method, comprising:

exposing a plurality of primary nucleic acid probes to a sample; and

exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample, wherein at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

93. The method of claim 92, further comprising:

creating codewords based on fluorescence of the secondary nucleic acid probes.

94. The method of claim 93, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

95. A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

96. The composition of claim 95, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

97. The composition of claim 96, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 3.

98. The composition of any one of claims 96 or 97, wherein the code is a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, or a Hamming(127, 120) code.

99. The composition of any one of claims 95-98, wherein the code is a SECDED code.

100. The composition of claim 99, wherein the code is a SECDED(8,4) code.

101. The composition of claim 99, wherein the code is a SECDED(16,4) code.

102. The composition of claim 99, wherein the code is a SECDED(16, 11) code, a SECDED (22, 16) code, a SECDED (39, 32) code, or a SECDED (72, 64) code.

103. The composition of any one of claims 99-102, wherein only codewords with a constant number of ones are used.

104. The composition of claim 99, wherein the code is MHD4 code.

105. The composition of claim 99, wherein the code is MHD2 code.

106. The composition of any one of claims 95-105, wherein the plurality of nucleic acid probes comprises no more than 100 possible read sequences.

107. The composition of any one of claims 95-106, wherein the plurality of nucleic acid probes comprises no more than 64 possible read sequences.

108. The composition of any one of claims 95-107, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

109. The composition of any one of claims 95-108, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

110. The composition of any one of claims 95-109, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

111. The composition of any one of claims 95-110, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

112. The composition of any one of claims 95-111, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

113. The composition of any one of claims 95-112, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

114. The composition of any one of claims 95-113, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

115. The composition of any one of claims 95-114, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

116. The composition of any one of claims 95-115, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

117. The composition of any one of claims 95-116, wherein at least some of the plurality of nucleic acid probes comprise DNA.

118. The composition of any one of claims 95-117, wherein at least some of the plurality of nucleic acid probes comprise RNA.

119. The composition of any one of claims 95-118, wherein at least some of the plurality of nucleic acid probes comprise PNA.

120. The composition of any one of claims 95-119, wherein at least some of the plurality of nucleic acid probes comprise LNA.

121. The composition of any one of claims 95-120, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

122. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

123. The method of claim 122, wherein the code space has a Hamming distance of at least 2.

124. The method of any one of claims 122 or 123, wherein the plurality of targets comprise RNA.

125. The method of any one of claims 122-124, wherein the plurality of targets comprise DNA.

126. A method, comprising:

exposing nucleic acid probes formed using the method of any one of claims 122-125 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets

within the sample.

127. The method of claim 126, wherein at least some of the plurality of secondary nucleic probes is able to bind to a read sequence of a nucleic acid probe.

128. The method of claim 127, wherein at least some of the plurality of secondary nucleic probes is able to specifically bind to only one read sequence.

129. The method of claim 125-128, wherein if no match is found between a valid codeword and the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

130. The method of claim 125-129, wherein at least some of the secondary nucleic acid probes each comprise a signaling entity.

131. The method of claim 130, comprising imaging the signaling entity to determine binding of the plurality of secondary nucleic probes within the sample.

132. The method of claim 131, comprising imaging the signaling entity using a super-resolution imaging technique.

133. The method of any one of claims 131 or 132, further comprising applying drift correction to the images.

134. The method of any one of claims 131-133, further comprising applying physical registration to the images.

135. The method of claim 125-134, comprising determining binding of the plurality of secondary nucleic probes within the sample at a resolution better than 100 nm.

136. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-

correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

137. The method of claim 136, where in the codewords comprise a number of positions that is less than the number of targets.

138. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

139. The method of claim 138, wherein the codewords form an error-checking space.

140. The method of any one of claims 138 or 139, wherein the codewords form an error-correcting code space.

141. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each

group of nucleic acid probes correspond to at least one common value of a position within the codewords.

142. The method of any one of claims 138-141, wherein the groups are applied to the sample sequentially using fluidic devices.

143. The method of any one of claims 138-142, wherein each group of nucleic acid probes contains a group identification sequence distinguishable from the group identification sequence of the other groups.

144. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code;

associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample;

forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences;

forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences;

exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample;

exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

145. The method of claim 144, wherein the wherein the read sequences are taken from a set of orthogonal sequences, which have a homology of less than 15 basepairs with

one another and with the nucleic acid species in a sample.

146. The method of any one of claims 144 or 145, comprising combining the images of the sample into a combined image.

147. The method of claim 146, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

148. The method of any one of claims 144-147, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

149. The method of any one of claims 144-148, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

150. The method of any one of claims 144-149, comprising determining the centroid using a maximum likelihood algorithm.

151. The method of any one of claims 144-150, comprising determining the centroid using a least squares algorithm.

152. The method of any one of claims 144-151, comprising determining the centroid using a Bayesian algorithm.

153 The method of any one of claims 144-152, comprising determining the centroid using a compressed sensing algorithm.

154. The method of any one of claims 144-153, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

155. The method of claim 154, comprising matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a

valid codeword that correspond to a nucleic acid target

156. The method of any one of claims 144-155, comprising imaging the sample using fluorescence microscopy.

157. The method of any one of claims 144-156, comprising imaging the sample at a resolution better than 500 nm.

158. The method of any one of claims 144-157, comprising imaging the sample at a resolution better than 300 nm.

159. The method of any one of claims 144-158, comprising imaging the sample at a resolution better than 100 nm.

160. The method of any one of claims 144-159, comprising imaging the sample at a resolution better than 80 nm.

161. The method of any one of claims 144-160, comprising imaging the sample at a resolution better than 50 nm.

162. The method of any one of claims 144-161, comprising imaging the sample at a resolution better than 20 nm.

163. The method of any one of claims 144-162, comprising imaging the sample using a super-resolution fluorescence imaging technique.

164. The method of any one of claims 144-163, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

165. The method of any one of claims 144-164, where in the codewords comprise a number of positions that is less than the number of targets.

166. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets;

grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords;

exposing a sample to one of the probe pools;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

167. The method of claim 166, comprising combining the images of the sample into a combined image.

168. The method of claim 167, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

169. The method of claim 168, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

170. The method of claim 169, comprising matching the determined codeword to a valid codeword wherein, if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target.

171. The method of any one of claims 166-170, comprising imaging the sample using fluorescence microscopy.

172. The method of any one of claims 166-171, comprising imaging the sample at a resolution better than 300 nm.

173. The method of any one of claims 166-172, comprising imaging the sample at a resolution better than 100 nm.

174. The method of any one of claims 166-173, comprising imaging the sample using a super-resolution fluorescence imaging technique.

175. The method of any one of claims 166-174, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

176. The method of any one of claims 166-175, wherein the codewords comprise a number of positions that is less than the number of targets.

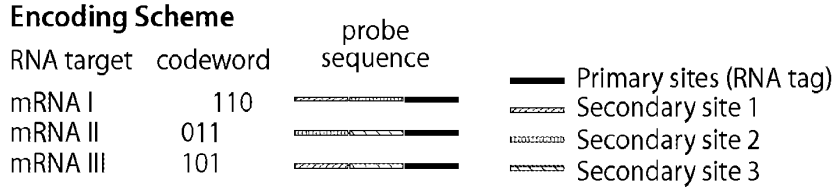


Fig. 1A

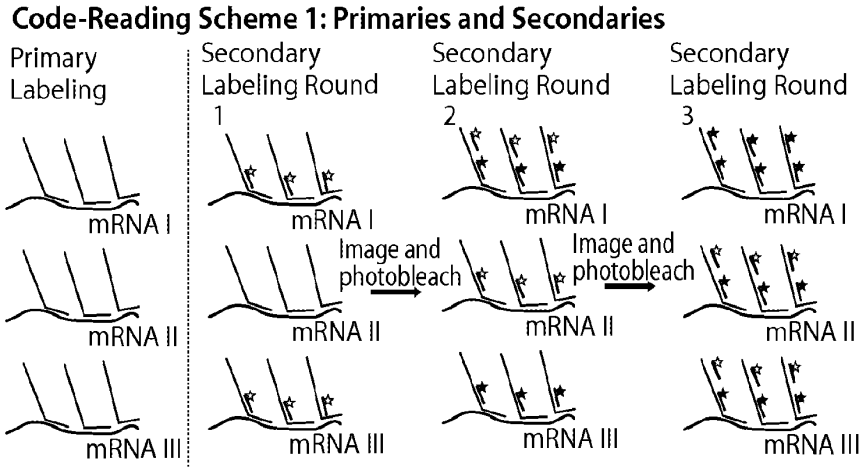


Fig. 1B

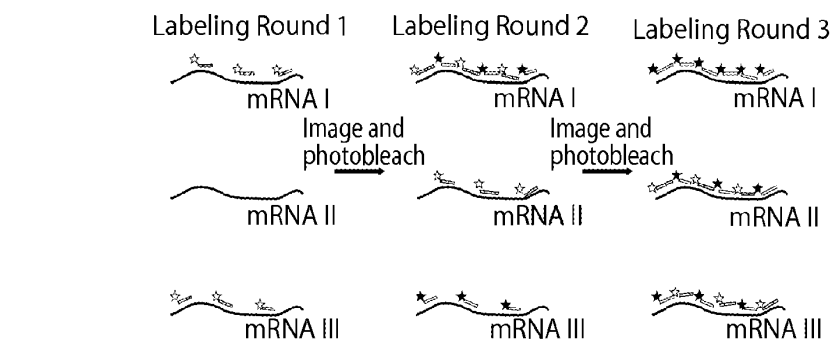


Fig. 1C

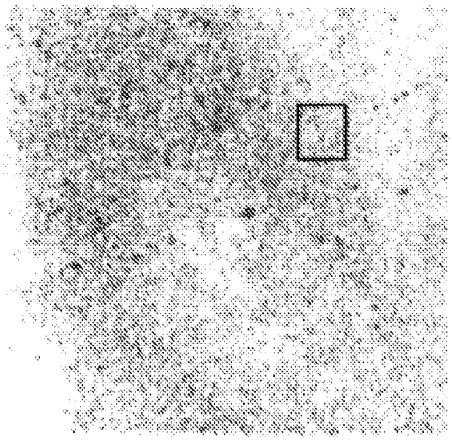


Fig. 2A

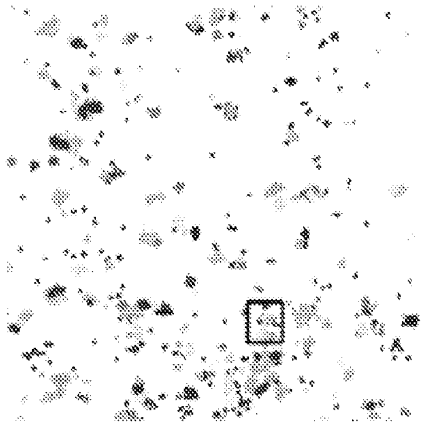


Fig. 2B



[0 1 0 1 1 1 0 0]

Fig. 2C

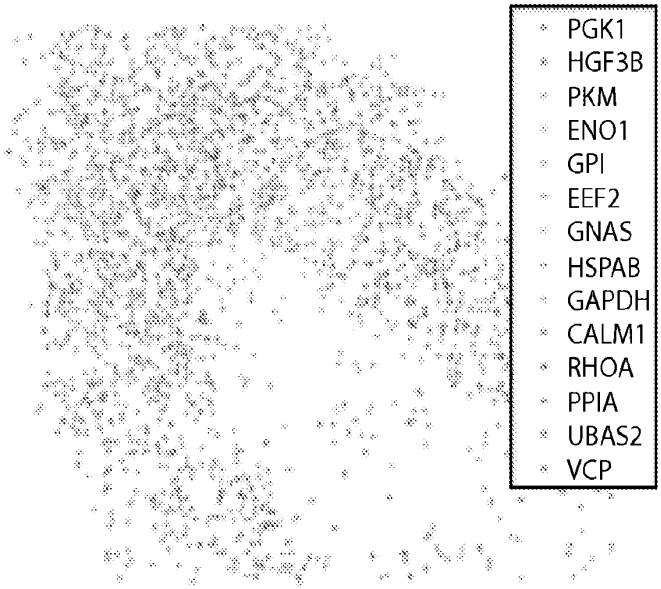

PGK1
HGF3B
PKM
ENO1
GPI
EEF2
GNAS
HSPAB
GAPDH
CALM1
RHOA
PPIA
UBAS2
VCP

Fig. 2D

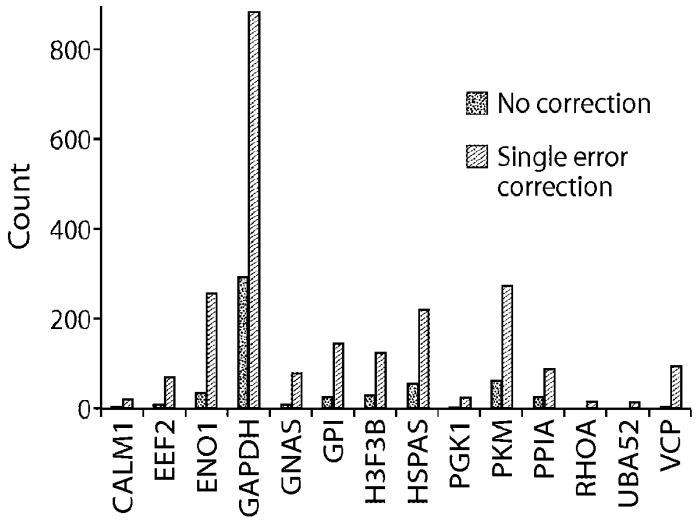

800
600
400
200
0
Count
No correction
Single error correction
CALM1
EEF2
ENO1
GAPDH
GNAS
GPI
H3F3B
HSPAS
PGK1
PKM
PPIA
RHOA
UBA52
VCP

Fig. 2E

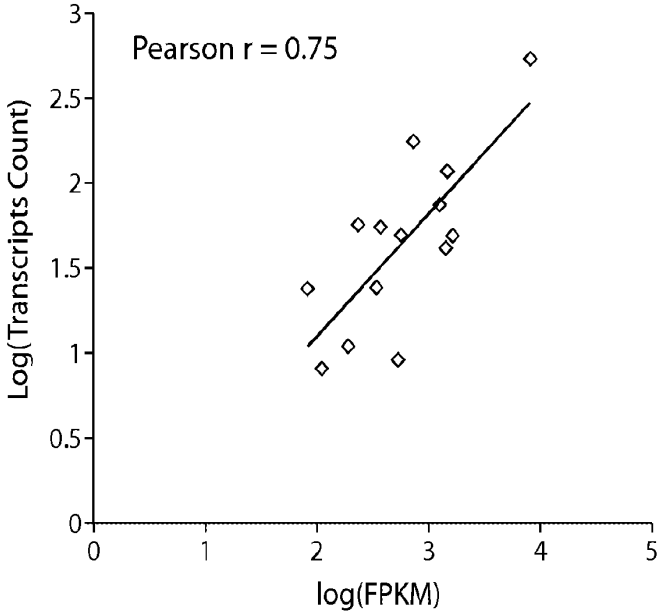

Pearson r = 0.75
Log(Transcripts Count)
log(FPKM)
0
0.5
1
1.5
2
2.5
3
0
1
2
3
4
5

Fig. 2F

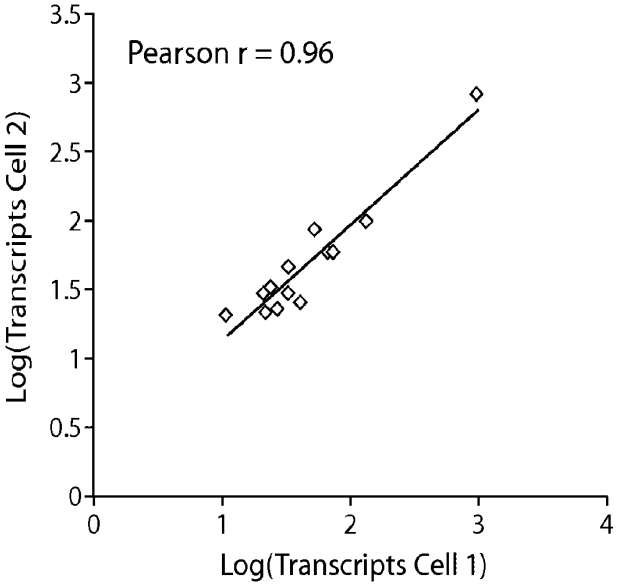

Pearson r = 0.96
Log(Transcripts Cell 2)
Log(Transcripts Cell 1)
0
0.5
1
1.5
2
2.5
3
3.5
0
1
2
3
4

Fig. 2G

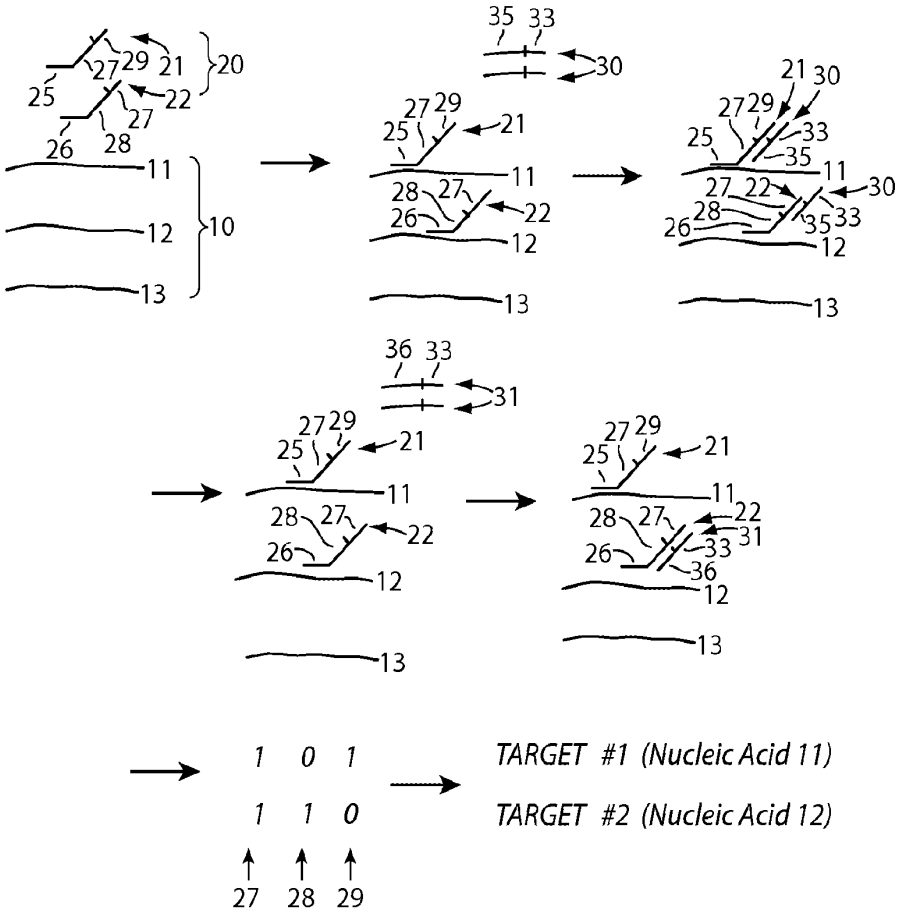

35 33
30
21 30
27 29
25
21
33
35
11
27 22
30
28
26
22
35 33
12
13
10
20
27 29 21
25
22
27
26 28
11
12
13
36 33
31
27 29
25
21
11
28 27
22
26
12
13
1 0 1
1 1 0
27 28 29
TARGET #1 (Nucleic Acid 11)
TARGET #2 (Nucleic Acid 12)

Fig. 3A

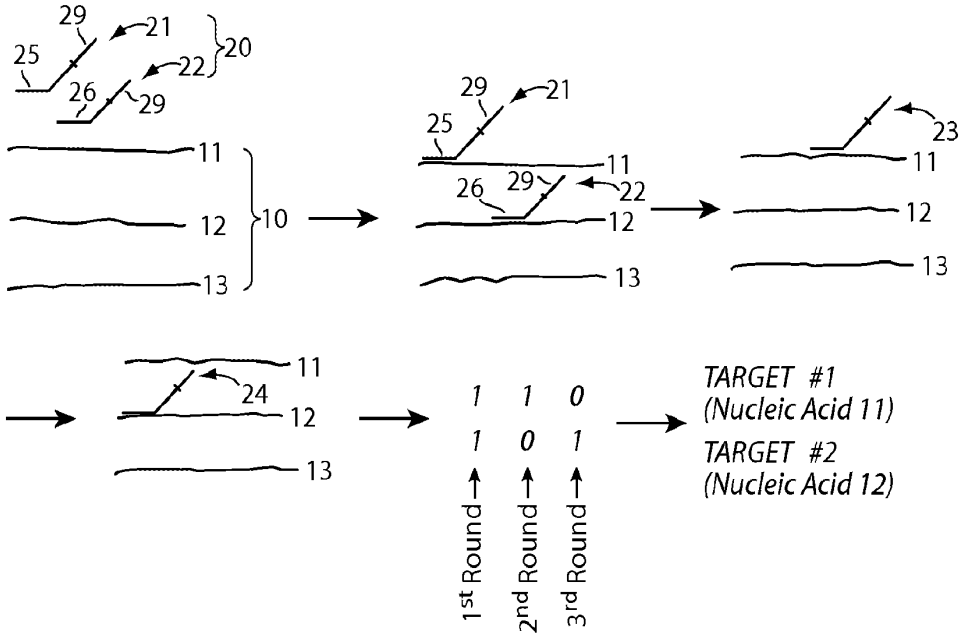


Fig. 3B

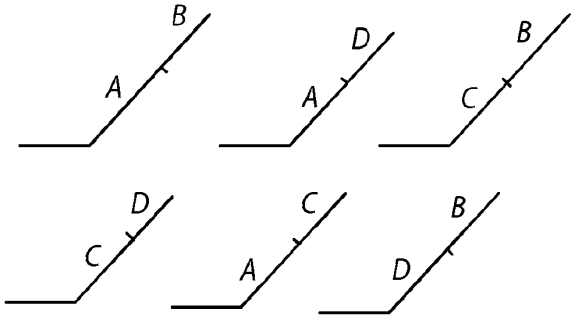


Fig. 4A

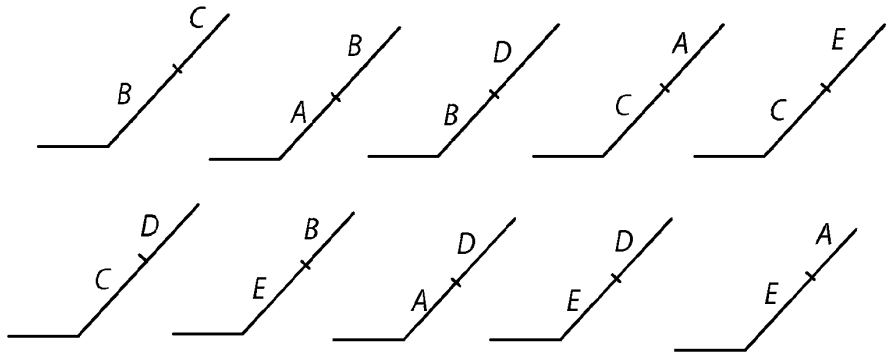


Fig. 4B

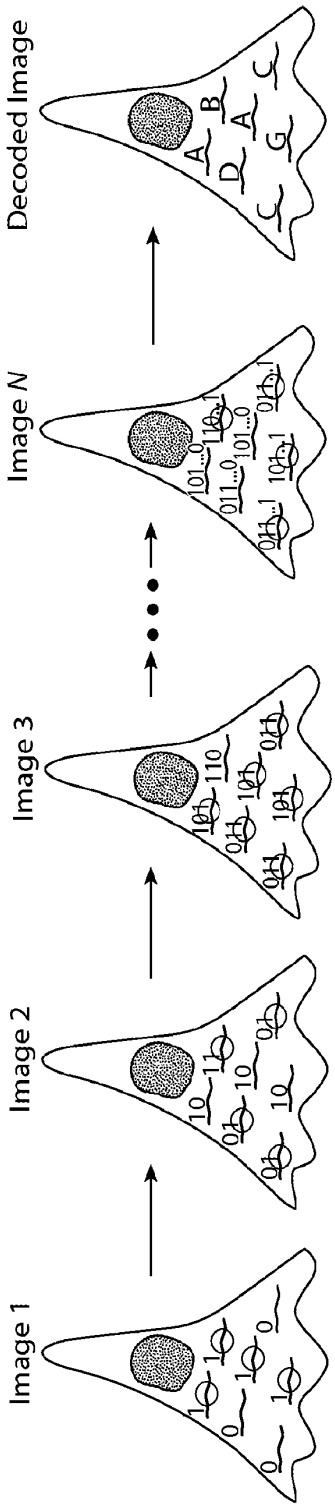


Fig. 5A

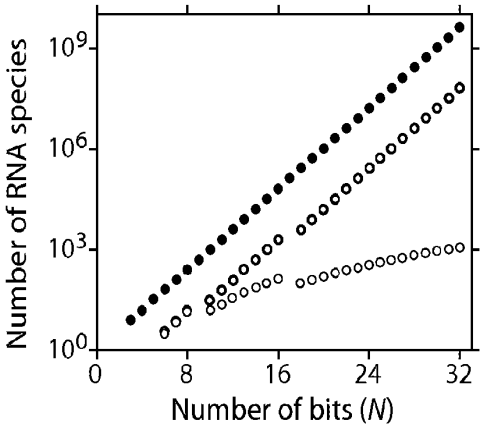


Fig. 5B

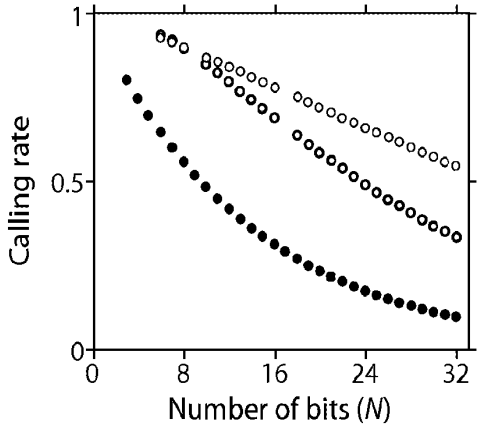


Fig. 5C

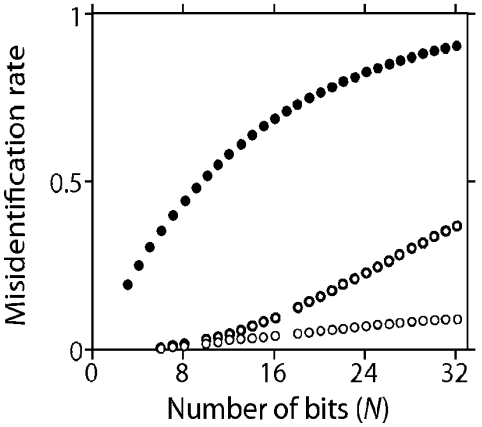


Fig. 5D

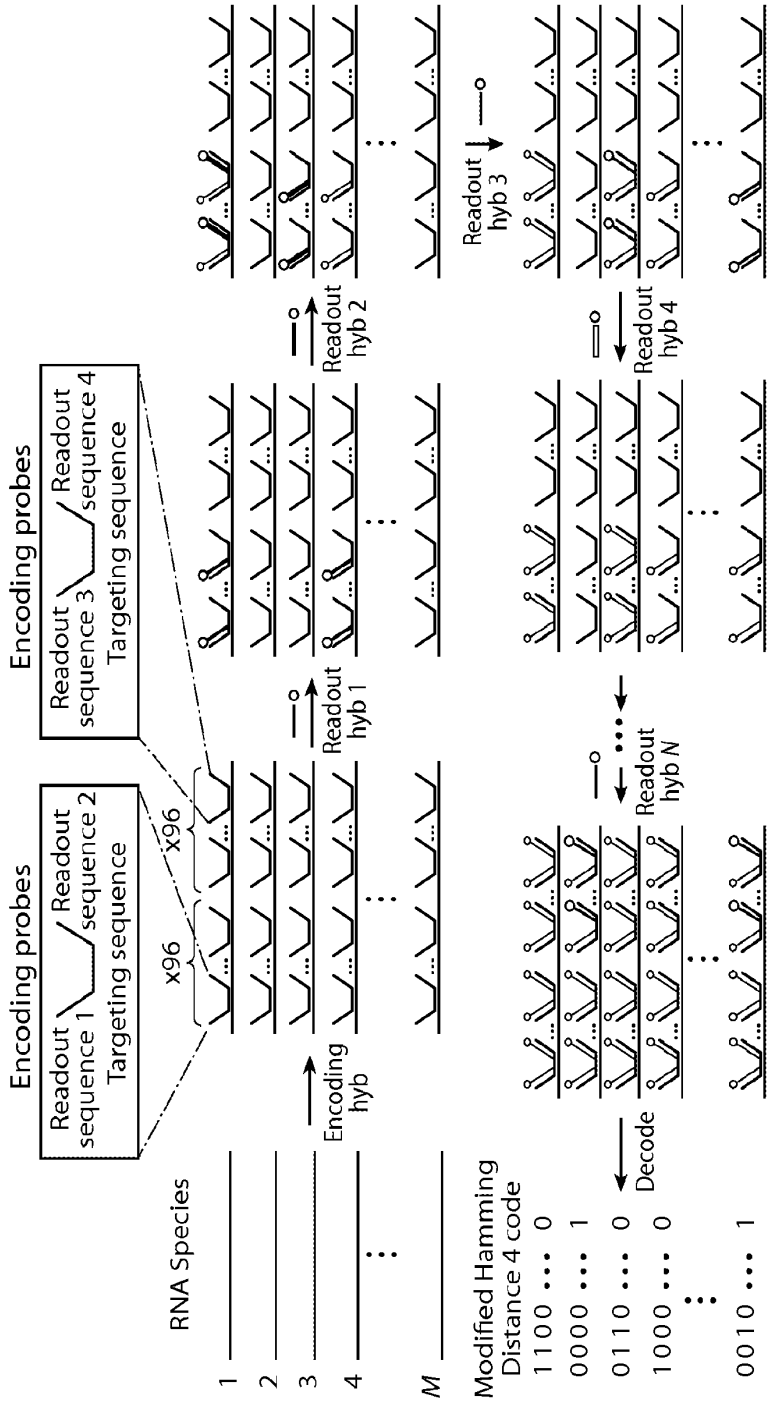

Encoding probes
Readout sequence 1
Readout sequence 2
Targeting sequence
Encoding probes
Readout sequence 3
Readout sequence 4
Targeting sequence
x96
x96
RNA Species
1
2
3
4
M
Encoding hyb
Readout hyb 1
Readout hyb 2
Readout hyb 3
Readout hyb 4
Readout hyb N
Modified Hamming Distance 4 code
1100 ... 0
0000 ... 1
0110 ... 0
1000 ... 0
0010 ... 1
Decode

Fig. 5E

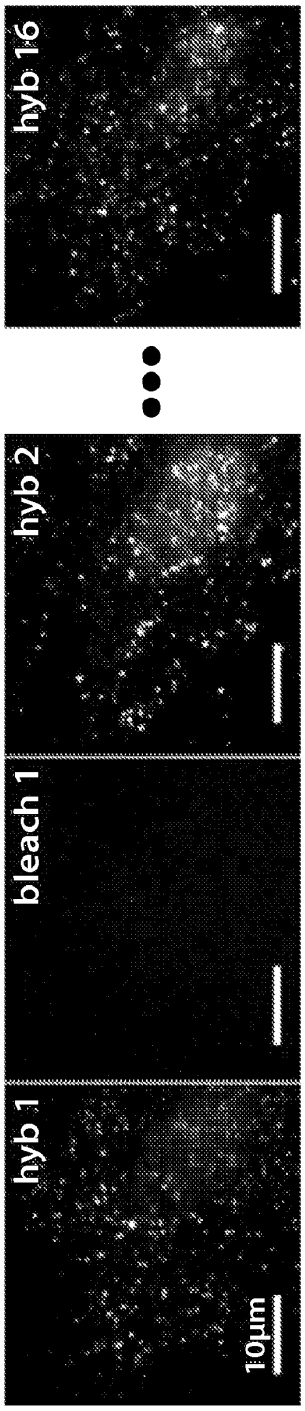


Fig. 6A

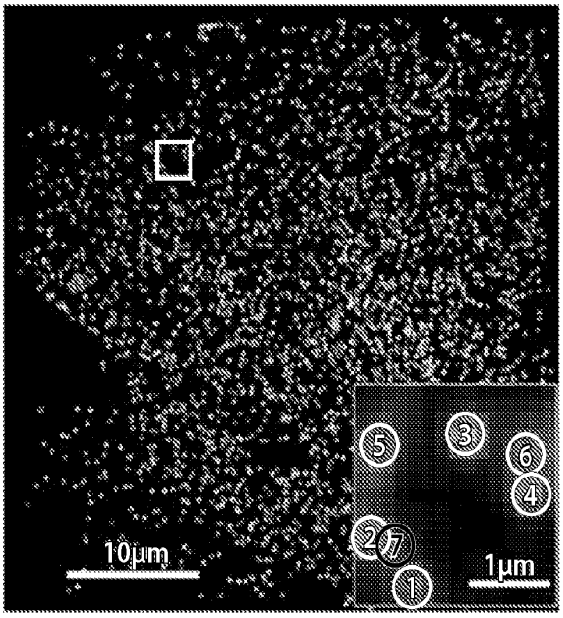

10µm
1µm
1
2
3
4
5
6
7

Fig. 6B

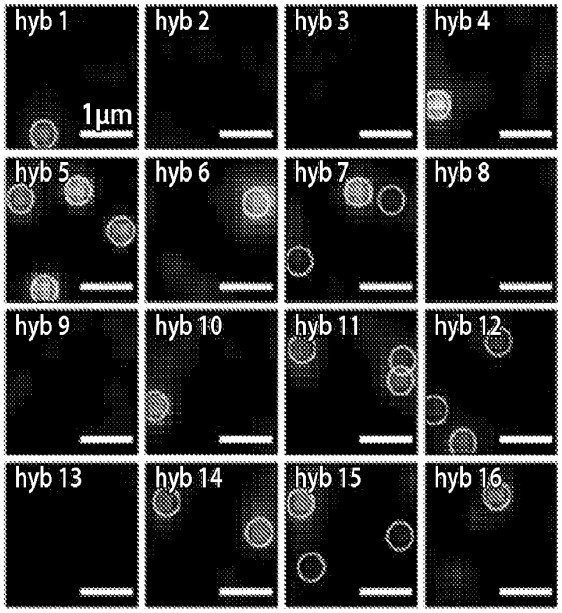

hyb 1
1µm
hyb 2
hyb 3
hyb 4
hyb 5
hyb 6
hyb 7
hyb 8
hyb 9
hyb 10
hyb 11
hyb 12
hyb 13
hyb 14
hyb 15
hyb 16

Fig. 6C

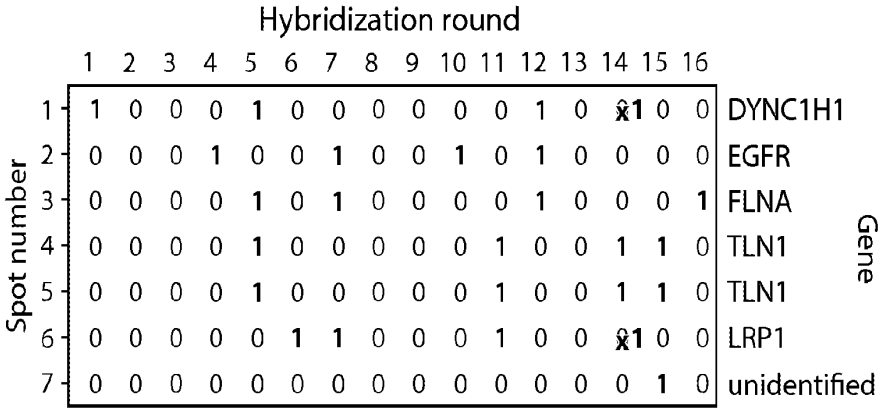

Hybridization round
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Spot number
Gene
1 1 0 0 0 1 0 0 0 0 0 0 1 0 x1 0 0 DYNC1H1
2 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 EGFR
3 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 FLNA
4 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 TLN1
5 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 TLN1
6 0 0 0 0 0 1 1 0 0 0 1 0 0 x1 0 0 LRP1
7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 unidentified

Fig. 6D

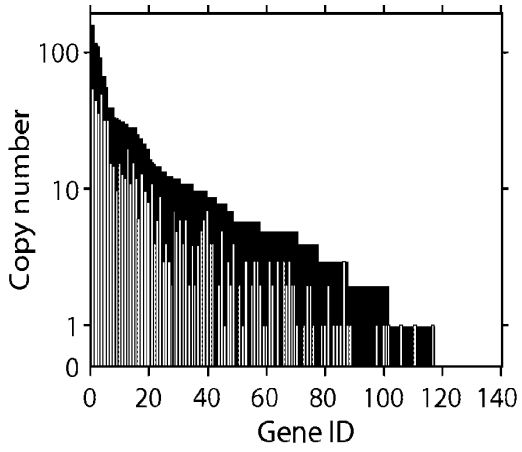

Copy number
100
10
1
0
0 20 40 60 80 100 120 140
Gene ID

Fig. 6E

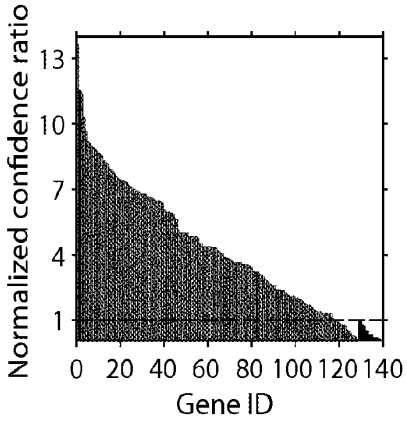


Fig. 6F

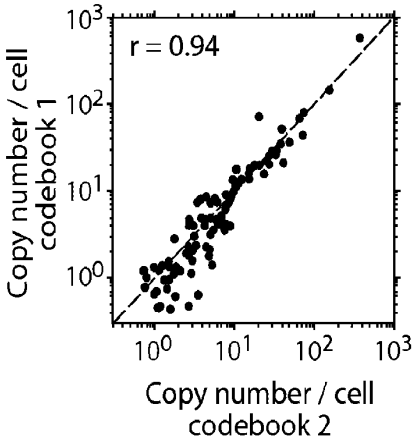


Fig. 6G

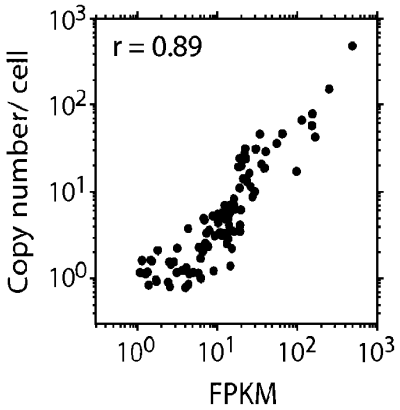


Fig. 6H

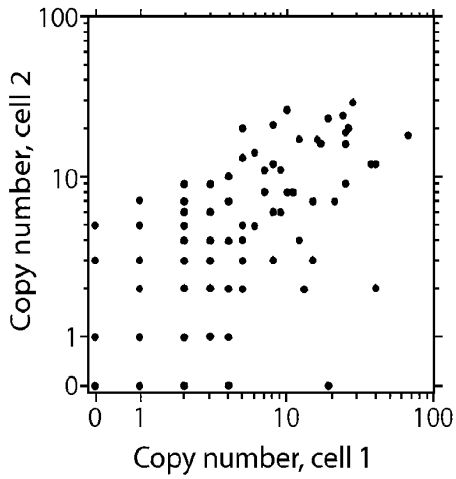

100
10
1
0
0
1
10
100
Copy number, cell 2
Copy number, cell 1

Fig. 7A

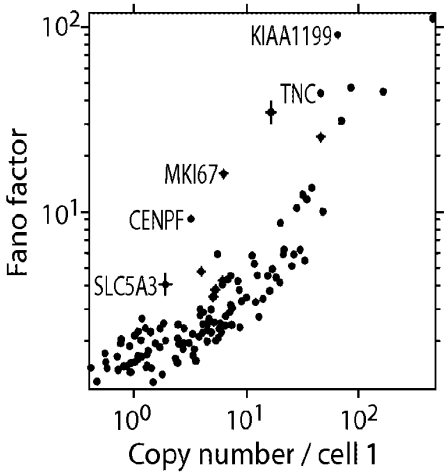

10^2
10^1
10^0
10^1
10^2
KIAA1199
TNC
MKI67
CENPF
SLC5A3
Fano factor
Copy number / cell 1

Fig. 7B

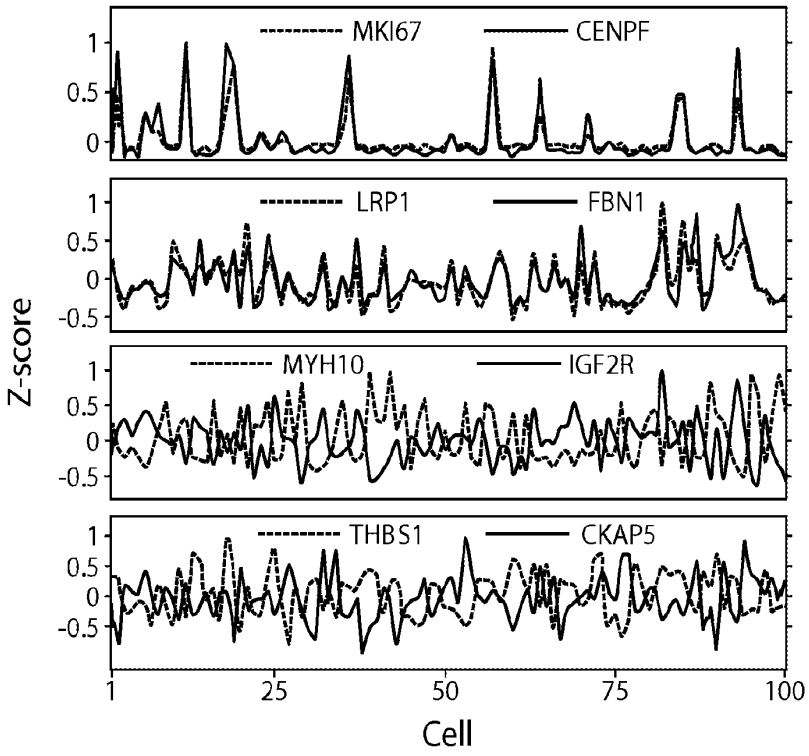

MKI67
CENPF
LRP1
FBN1
MYH10
IGF2R
THBS1
CKAP5
Z-score
1
0.5
0
-0.5
1
25
50
75
100
Cell

Fig. 7C

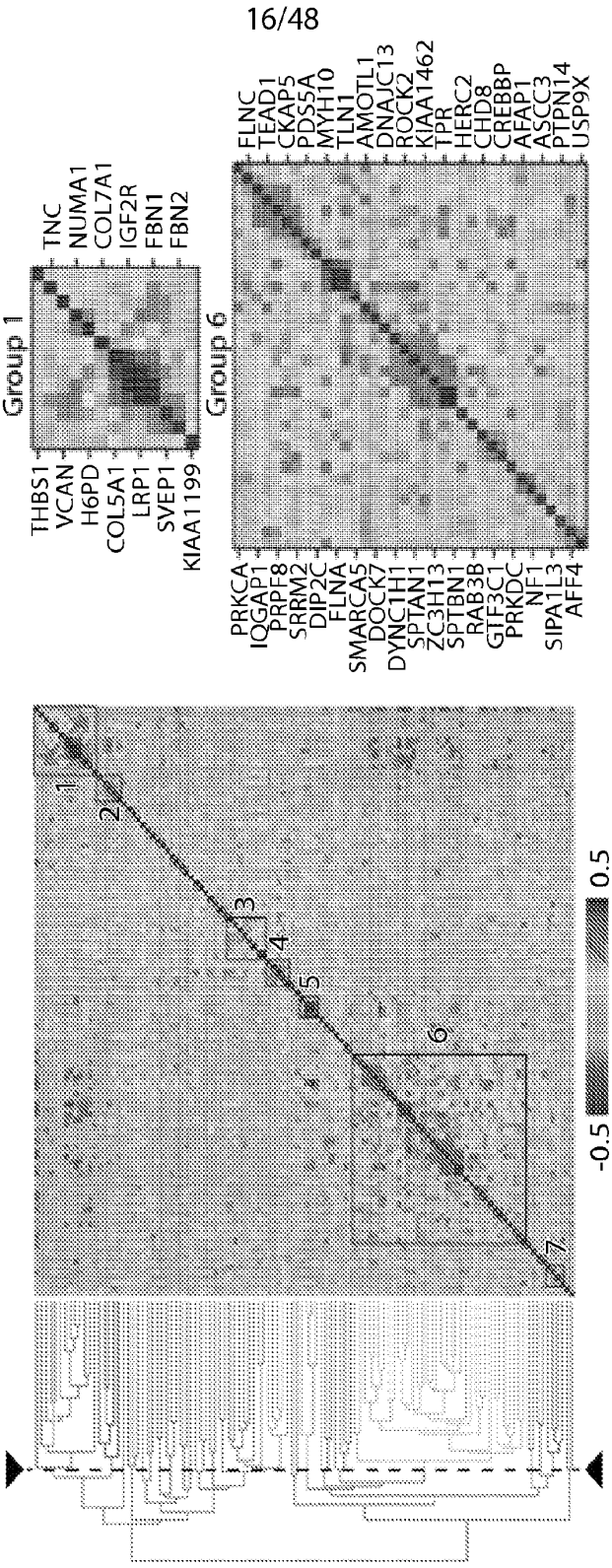

Group 1
THBS1
VCAN
H6PD
COL5A1
LRP1
SVEP1
KIAA1199
TNC
NUMA1
COL7A1
IGF2R
FBN1
FBN2
Group 6
PRKCA
IQGAP1
PRPF8
SRRM2
DIP2C
FLNA
SMARCA5
DOCK7
DYNC1H1
SPTAN1
ZC3H13
SPTBN1
RAB3B
GTF3C1
PRKDC
NF1
SIPA1L3
AFF4
FLNC
TEAD1
CKAP5
PDS5A
MYH10
TLN1
AMOTL1
DNAJC13
ROCK2
KIAA1462
TPR
HERC2
CHD8
CREBBP
AFAP1
ASCC3
PTPN14
USP9X
1
2
3
4
5
6
7
0.5
-0.5

Fig. 7D

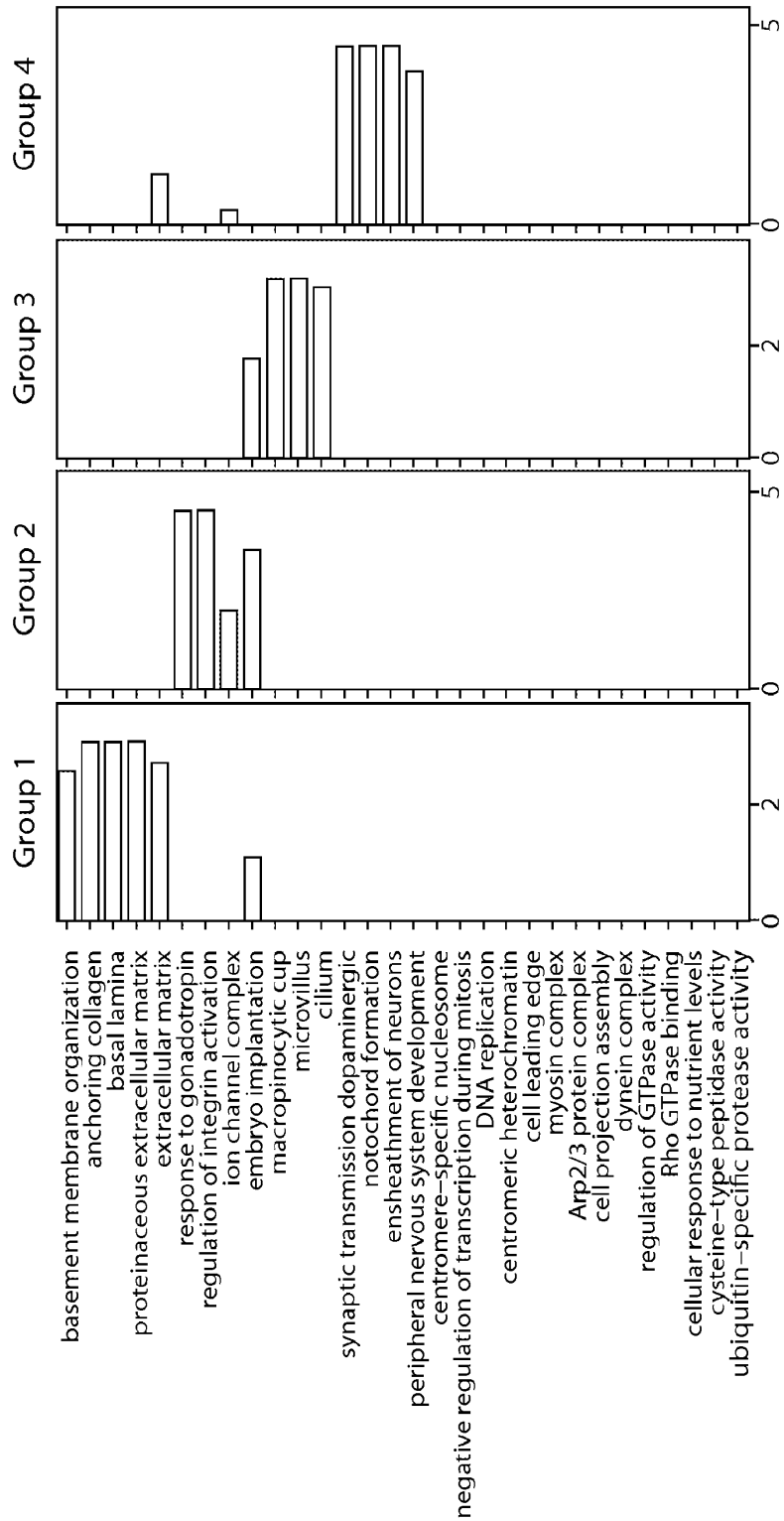

Group 1
Group 2
Group 3
Group 4
0
2
0
5
0
2
0
5
Enrichment (log2)
basement membrane organization
anchoring collagen
basal lamina
proteinaceous extracellular matrix
extracellular matrix
response to gonadotropin
regulation of integrin activation
ion channel complex
embryo implantation
macropinocytic cup
microvillus
cilium
synaptic transmission dopaminergic
notochord formation
ensheathment of neurons
peripheral nervous system development
centromere−specific nucleosome
negative regulation of transcription during mitosis
DNA replication
centromeric heterochromatin
cell leading edge
myosin complex
Arp2/3 protein complex
cell projection assembly
dynein complex
regulation of GTPase activity
Rho GTPase binding
cellular response to nutrient levels
cysteine−type peptidase activity
ubiquitin−specific protease activity

Fig. 7E

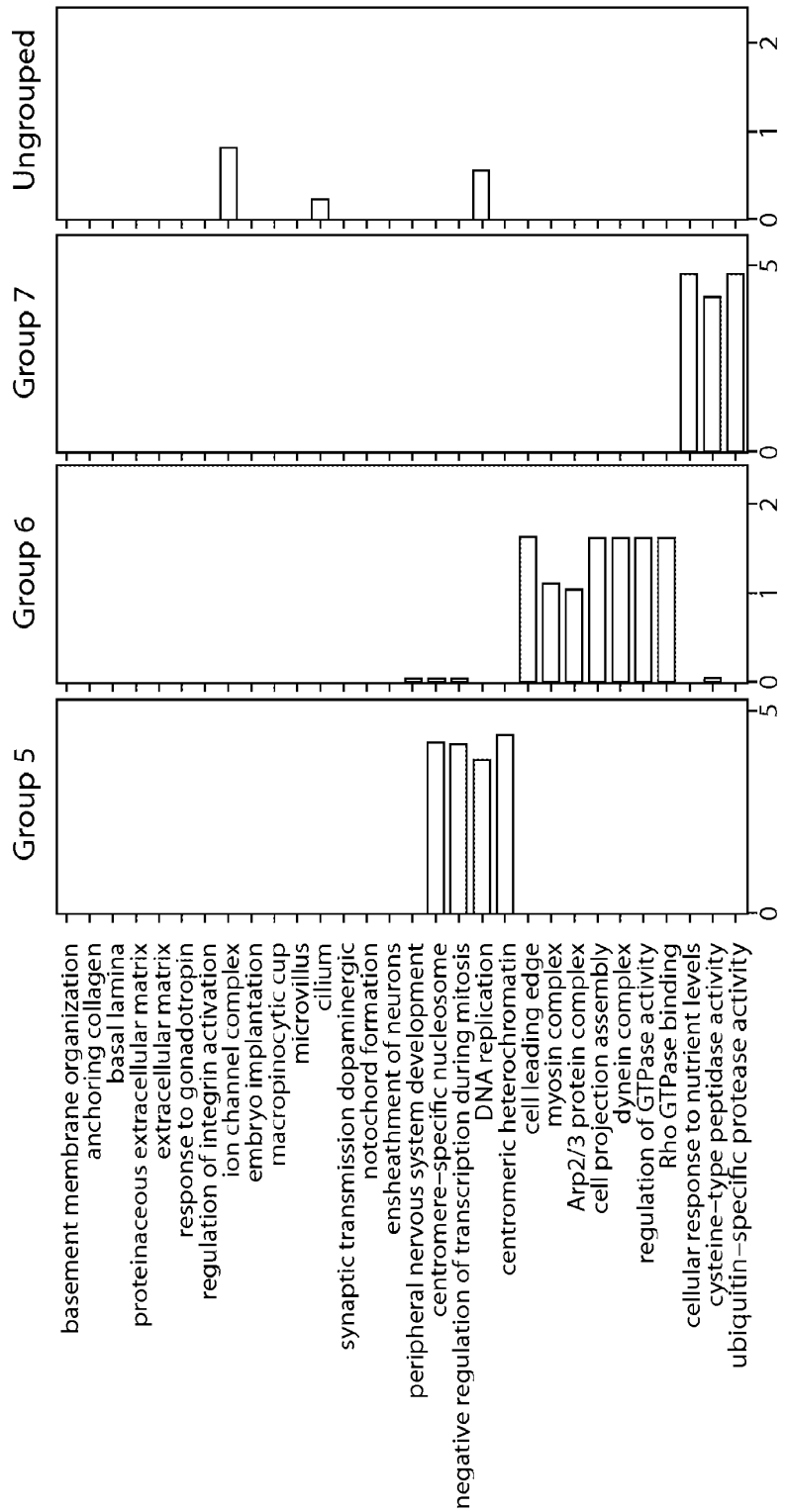


Fig. 7E Continued

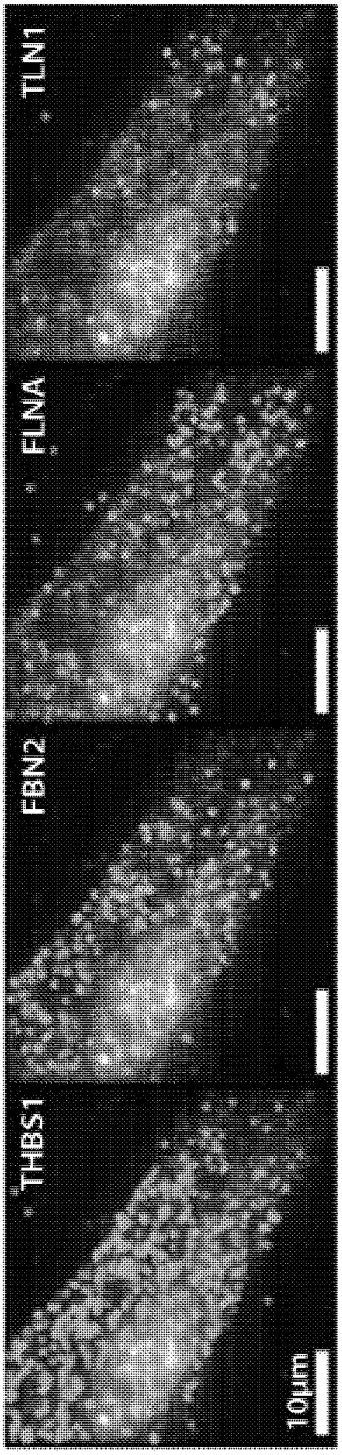


Fig. 8A

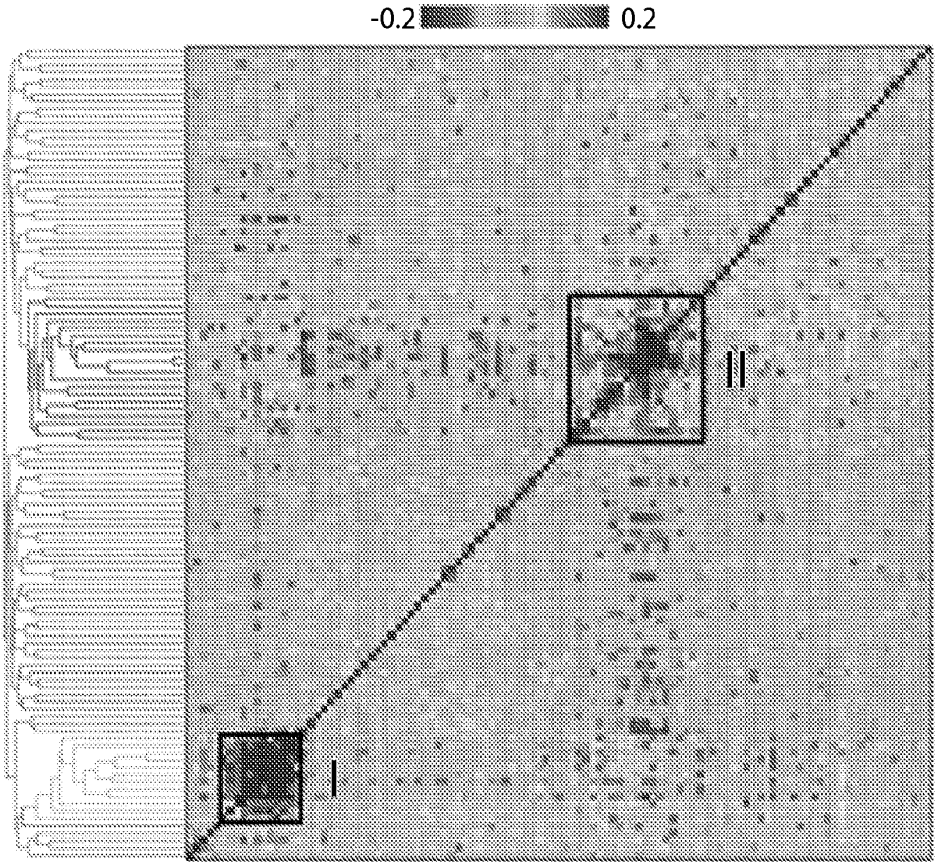


Fig. 8B

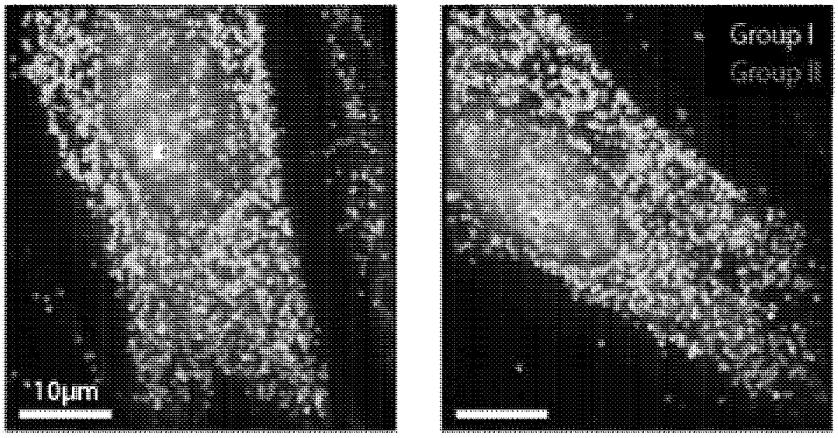


Fig. 8C

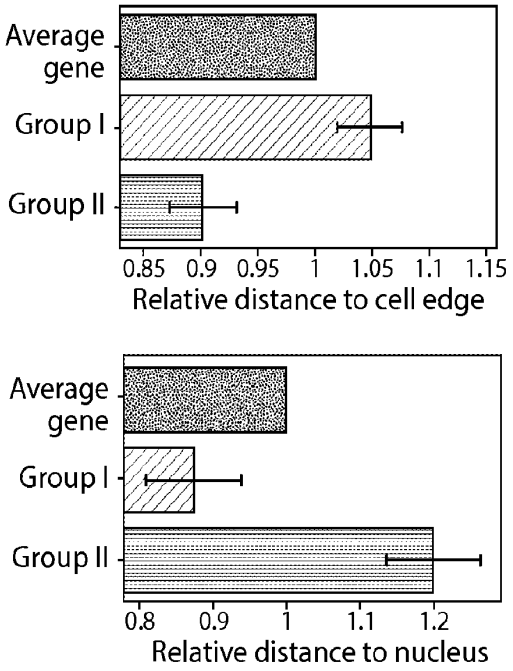

Average gene
Group I
Group II
0.85 0.9 0.95 1 1.05 1.1 1.15
Relative distance to cell edge
Average gene
Group I
Group II
0.8 0.9 1 1.1 1.2
Relative distance to nucleus

Fig. 8D

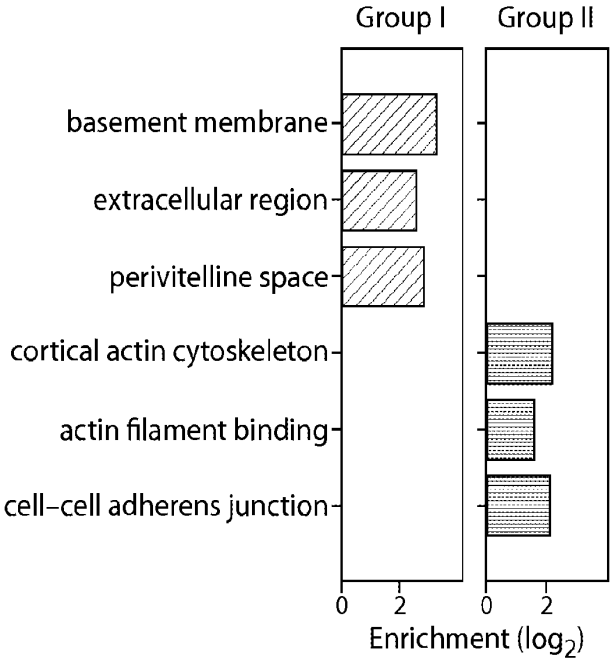

Group I
Group II
basement membrane
extracellular region
perivitelline space
cortical actin cytoskeleton
actin filament binding
cell–cell adherens junction
0 2
0 2
Enrichment (log_2)

Fig. 8E

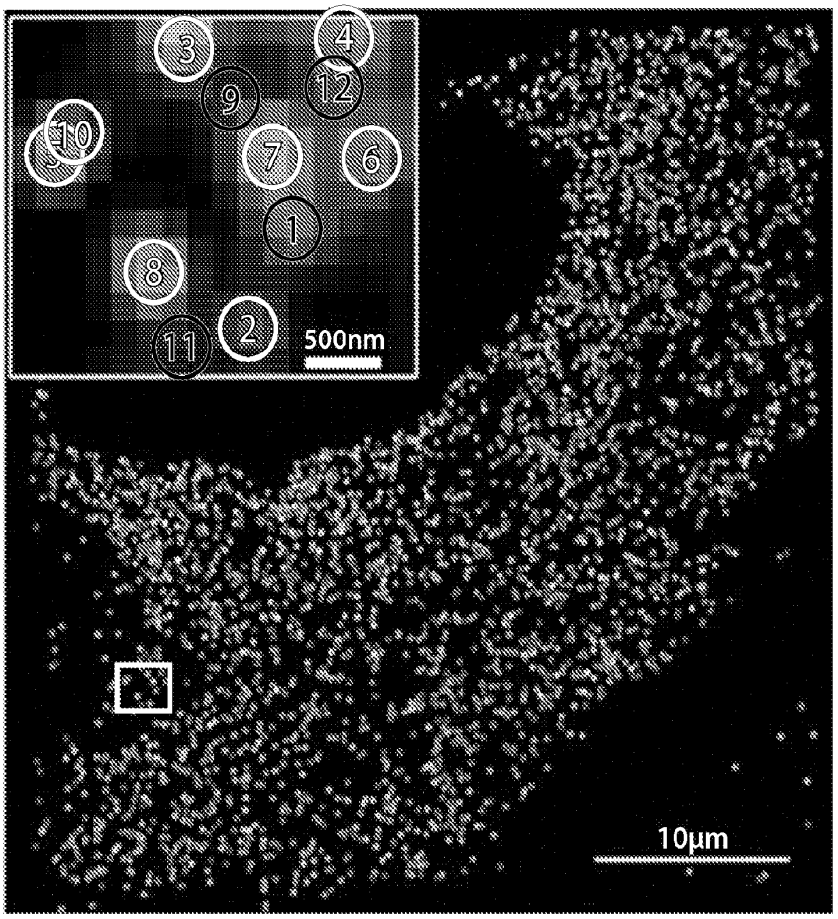


Fig. 9A

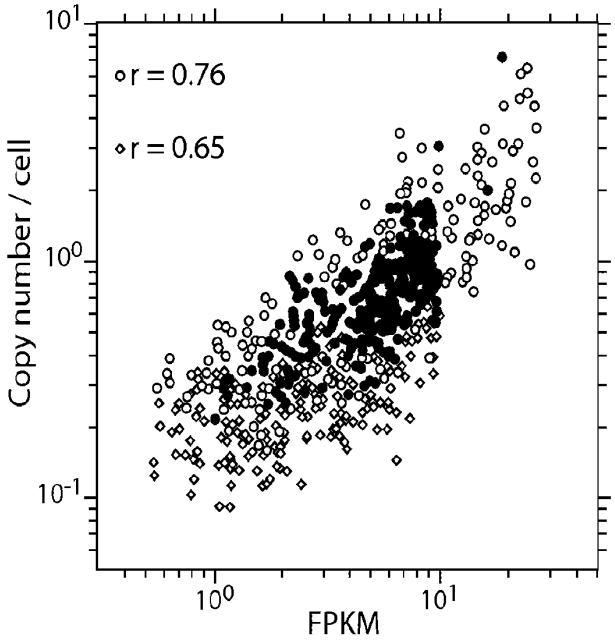

r = 0.76
r = 0.65
Copy number / cell
FPKM

Fig. 9B

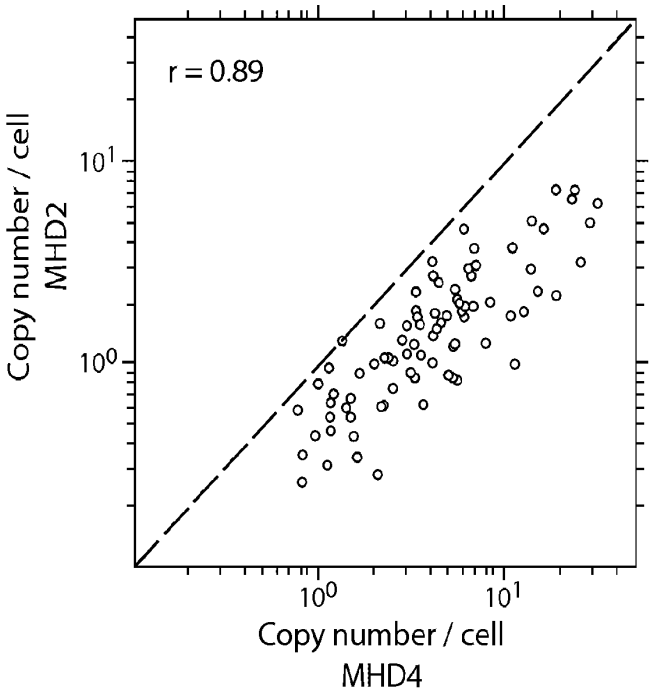

r = 0.89
Copy number / cell
MHD2
Copy number / cell
MHD4

Fig. 9C

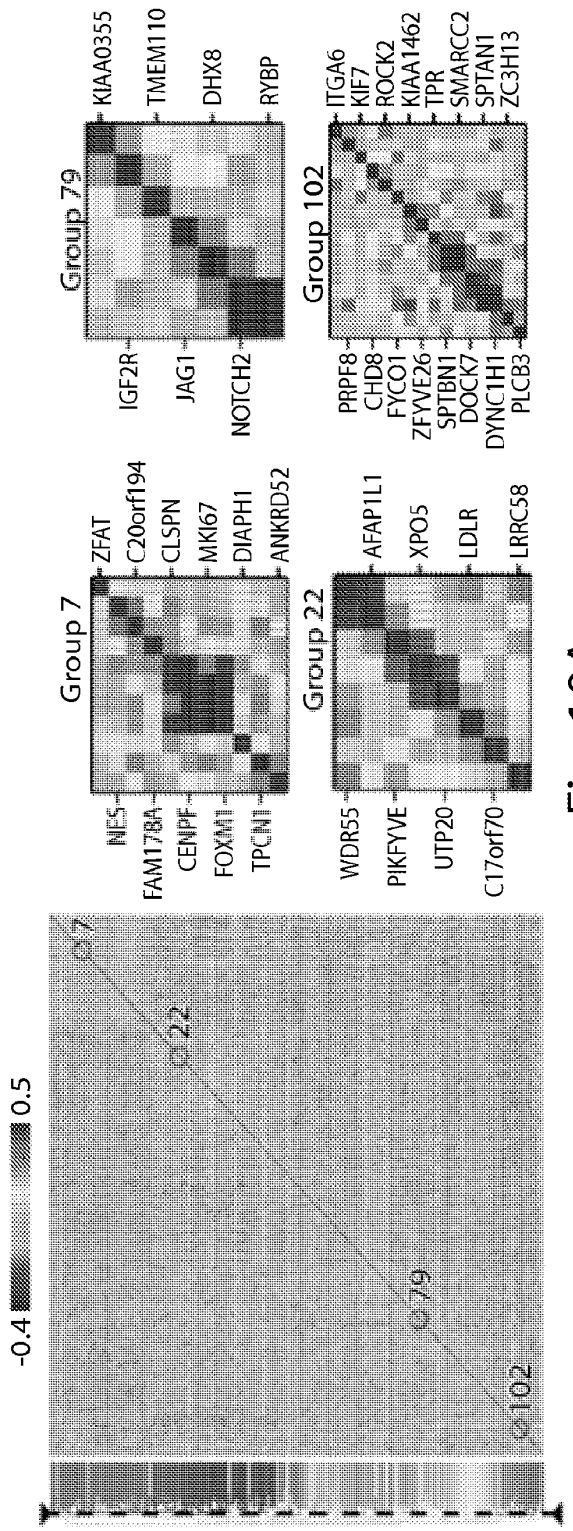

-0.4
0.5
g7
g22
g79
g102
Group 7
NES
FAM178A
CENPF
FOXM1
TPCN1
ZFAT
C20orf194
CLSPN
MKI67
DIAPH1
ANKRD52
Group 22
WDR55
PIKFYVE
UTP20
C17orf70
AFAP1L1
XPO5
LDLR
LRRC58
Group 79
IGF2R
JAG1
NOTCH2
KIAA0355
TMEM110
DHX8
RYBP
Group 102
PRPF8
CHD8
FYCO1
ZFYVE26
SPTBN1
DOCK7
DYNC1H1
PLCB3
ITGA6
KIF7
ROCK2
KIAA1462
TPR
SMARCC2
SPTAN1
ZC3H13

Fig. 10A

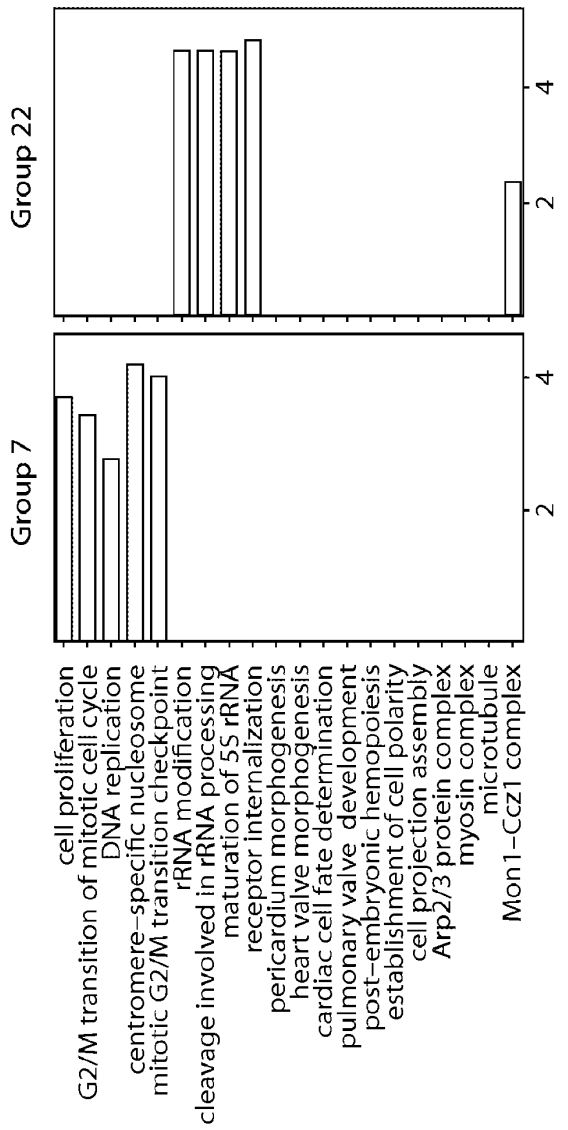

Group 7
Group 22
cell proliferation
G2/M transition of mitotic cell cycle
DNA replication
centromere−specific nucleosome
mitotic G2/M transition checkpoint
rRNA modification
cleavage involved in rRNA processing
maturation of 5S rRNA
receptor internalization
pericardium morphogenesis
heart valve morphogenesis
cardiac cell fate determination
pulmonary valve development
post−embryonic hemopoiesis
establishment of cell polarity
cell projection assembly
Arp2/3 protein complex
myosin complex
microtubule
Mon1−Ccz1 complex
2
4
2
4
Enrichment(log2)

Fig. 10B

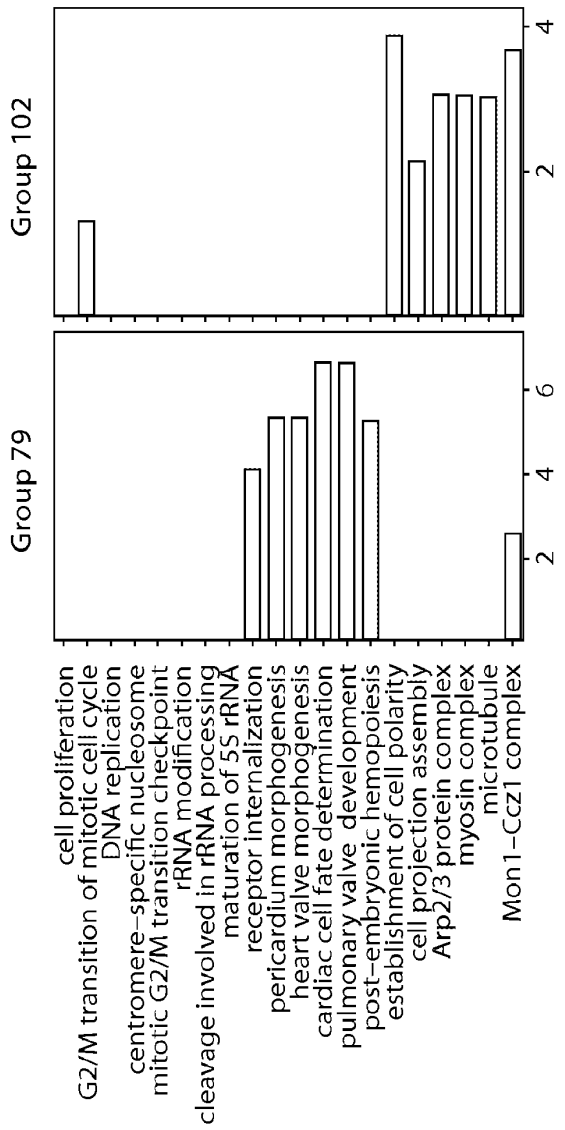


Fig. 10B Continued

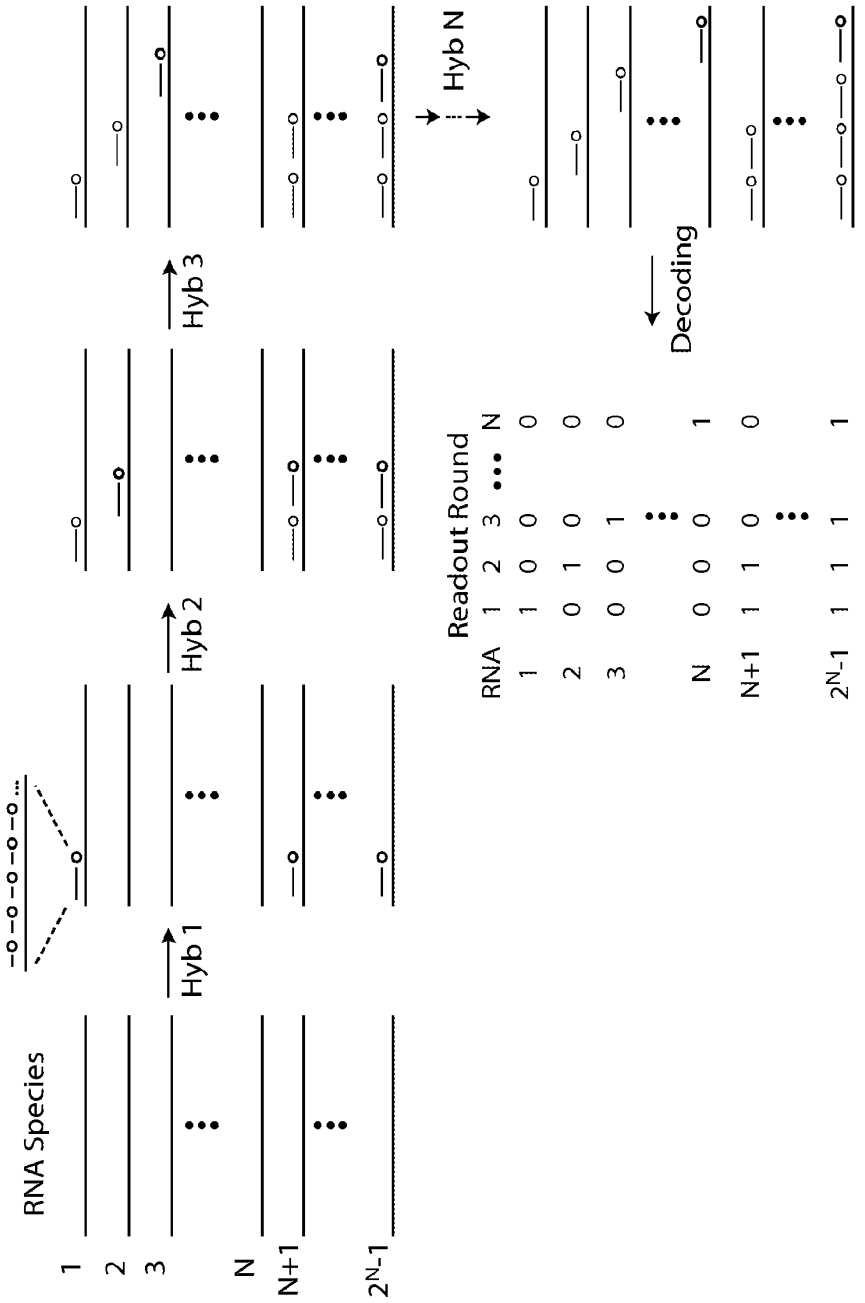

RNA Species
1
2
3
N
N+1
2N-1
Hyb 1
Hyb 2
Hyb 3
Hyb N
Readout Round
RNA 1 2 3 ... N
1 1 0 0 0
2 0 1 0 0
3 0 0 1 0
N 0 0 0 1
N+1 1 1 0 0
2N-1 1 1 1 1
Decoding

Fig. 11

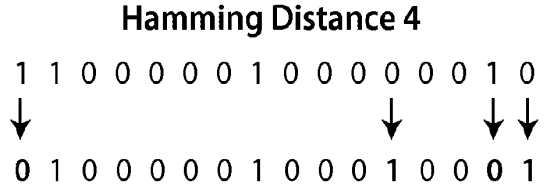


Fig. 12A

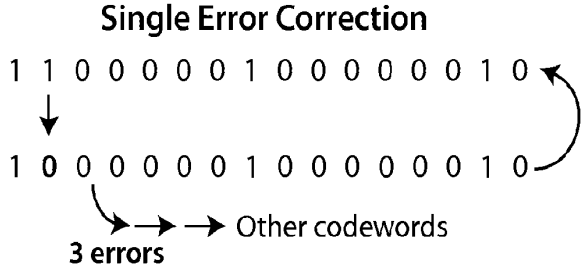


Fig. 12B

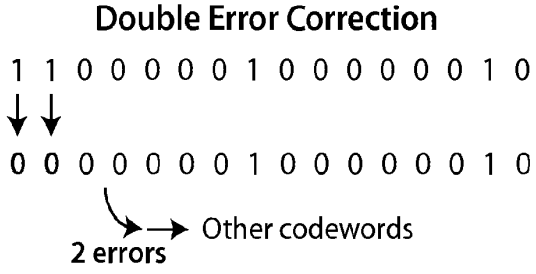


Fig. 12C

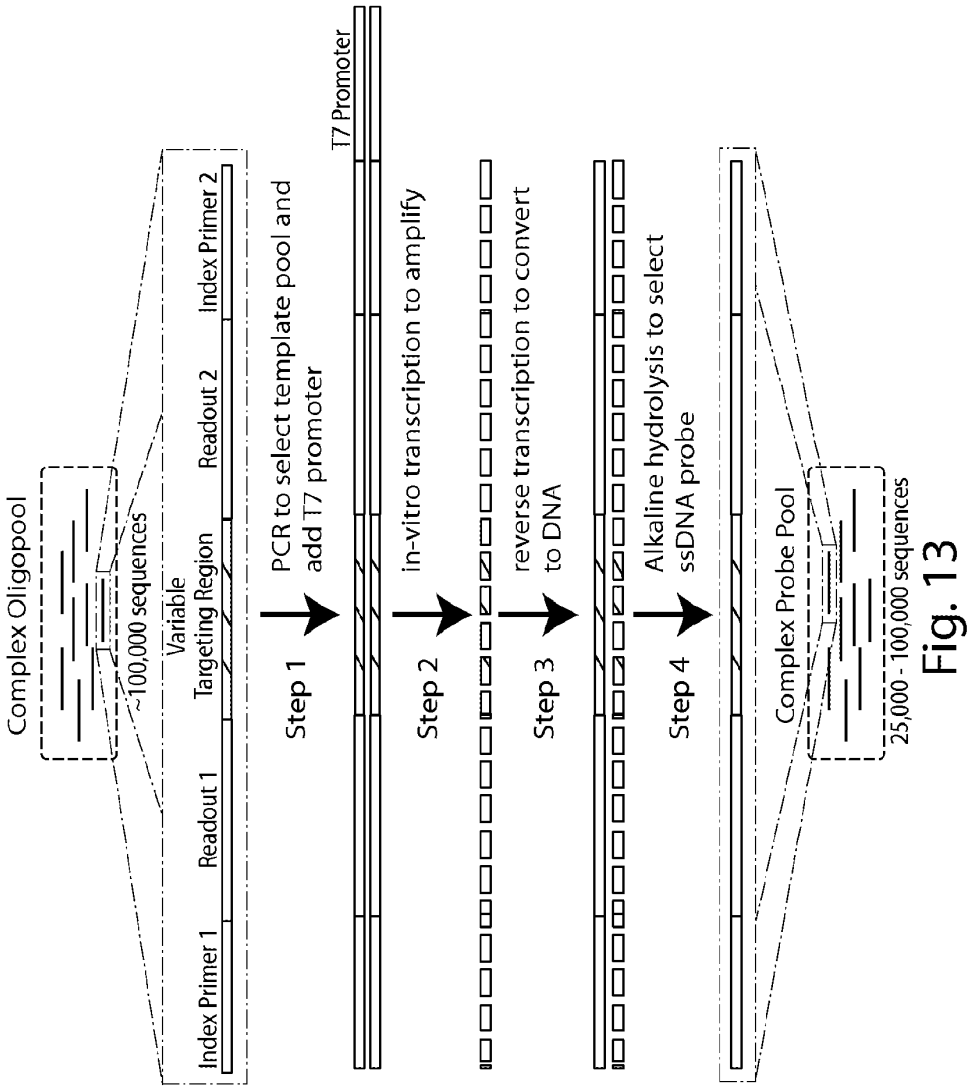

Complex Oligopool
~100,000 sequences
Variable
Index Primer 1
Readout 1
Targeting Region
Readout 2
Index Primer 2
Step 1
PCR to select template pool and add T7 promoter
T7 Promoter
Step 2
in-vitro transcription to amplify
Step 3
reverse transcription to convert to DNA
Step 4
Alkaline hydrolysis to select ssDNA probe
Complex Probe Pool
25,000 - 100,000 sequences

Fig. 13

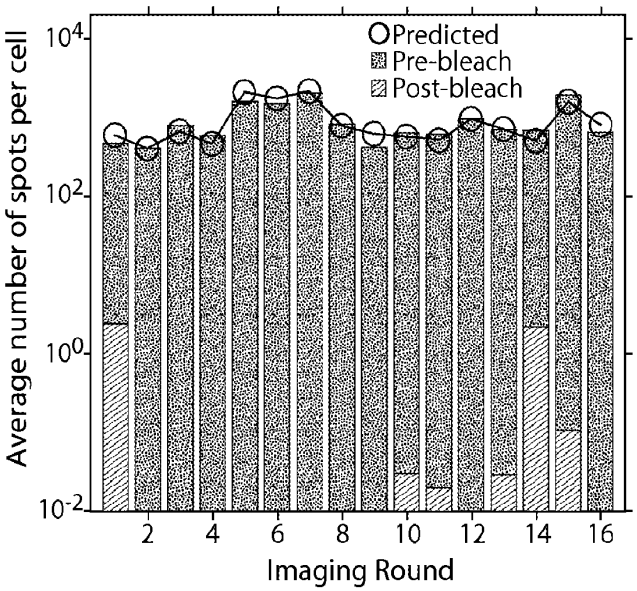


Fig. 14A

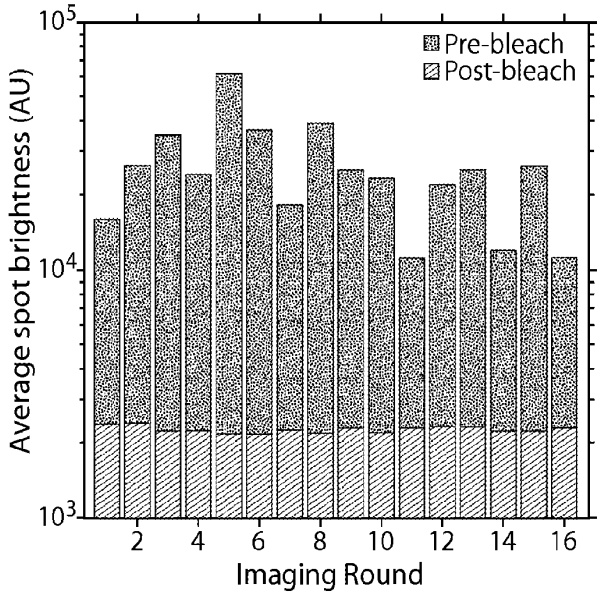


Fig. 14B

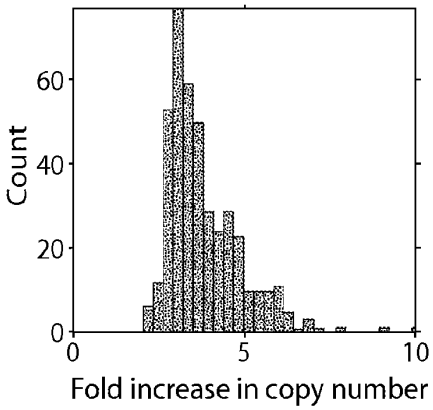


Fig. 15A

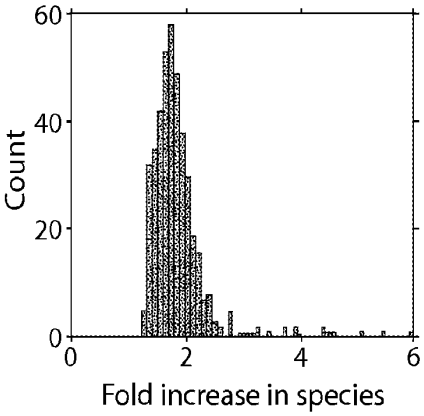


Fig. 15B

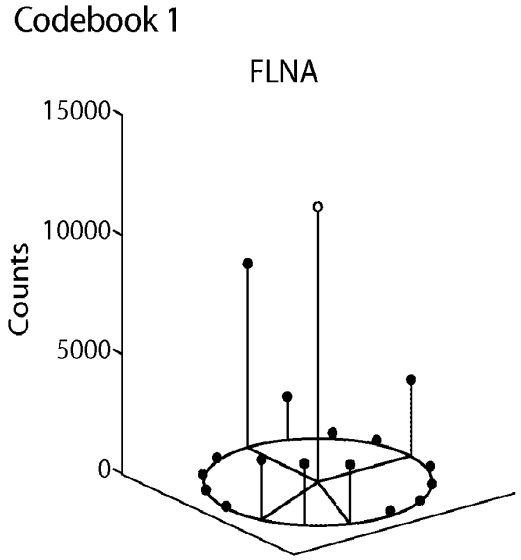


Fig. 16A

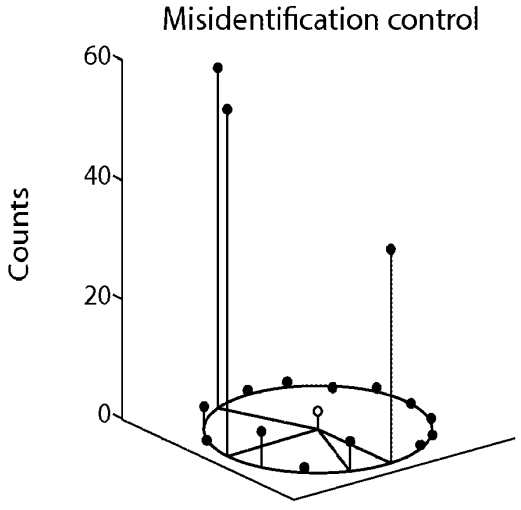


Fig. 16B

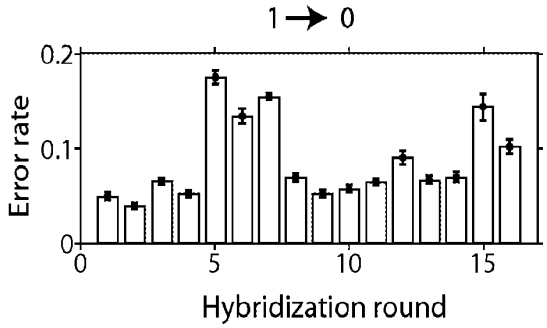


Fig. 16C

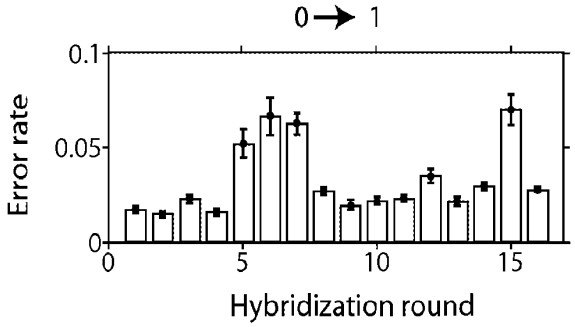


Fig. 16D

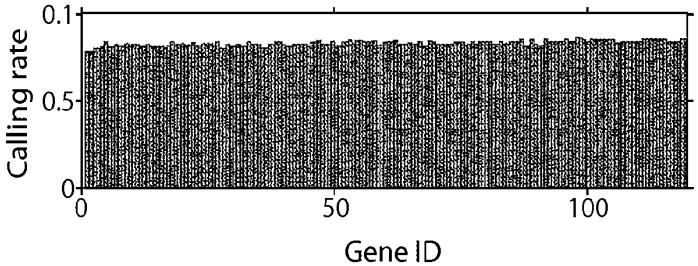


Fig. 16E

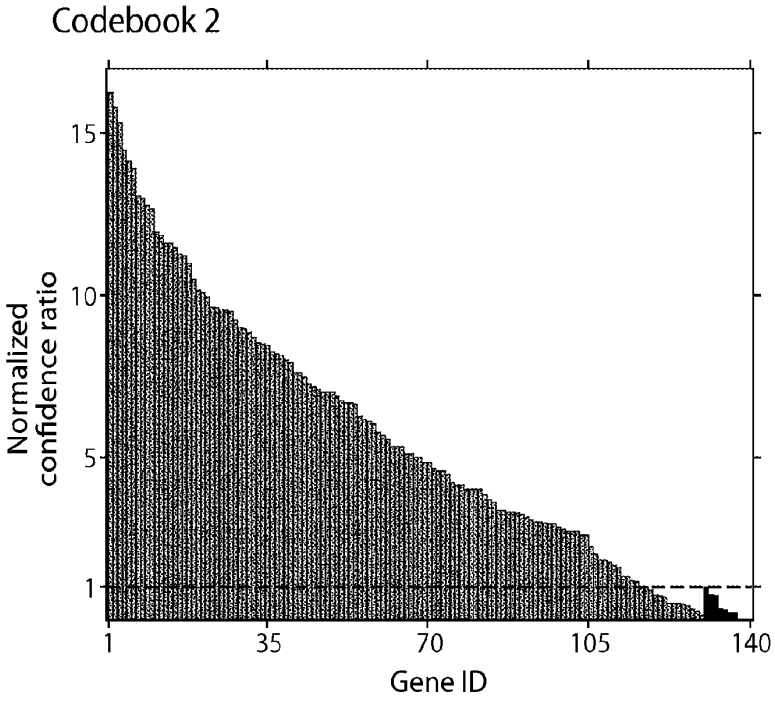


Fig. 17A

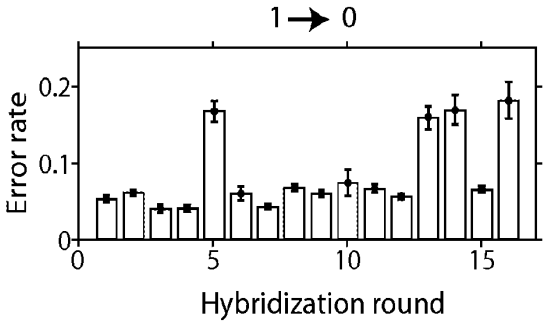


Fig. 17B

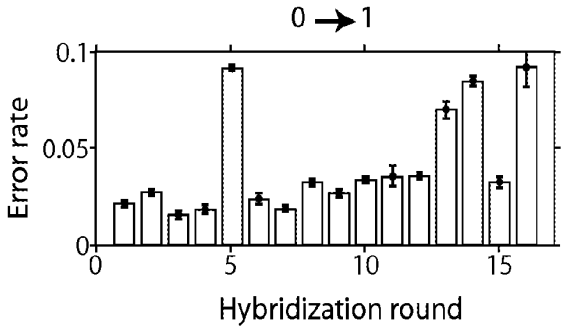


Fig. 17C

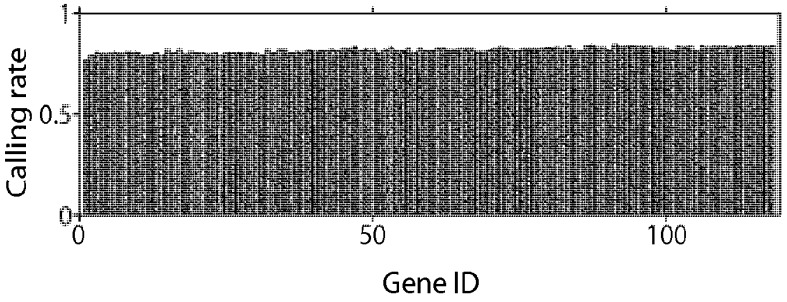


Fig. 17D

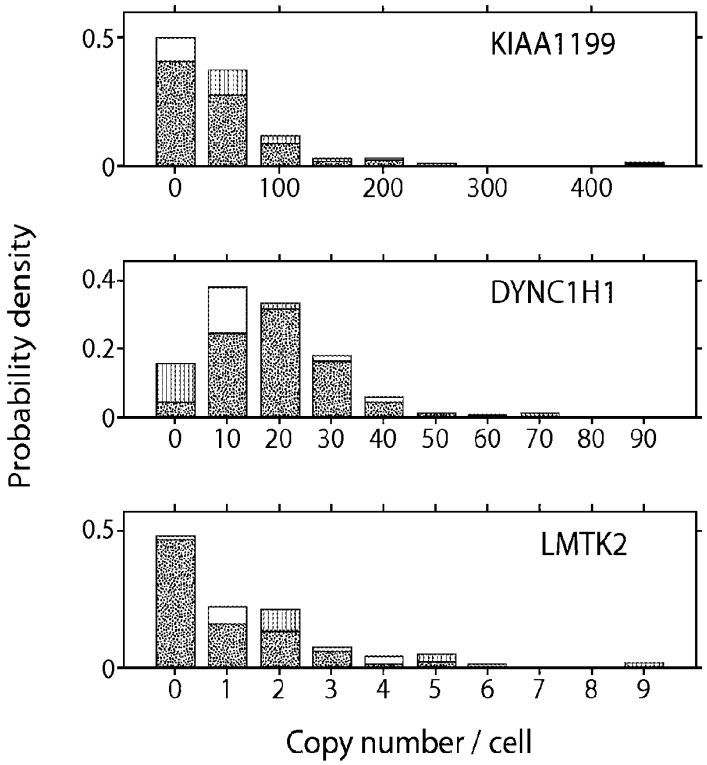

KIAA1199
DYNC1H1
LMTK2
Probability density
Copy number / cell

Fig. 18A

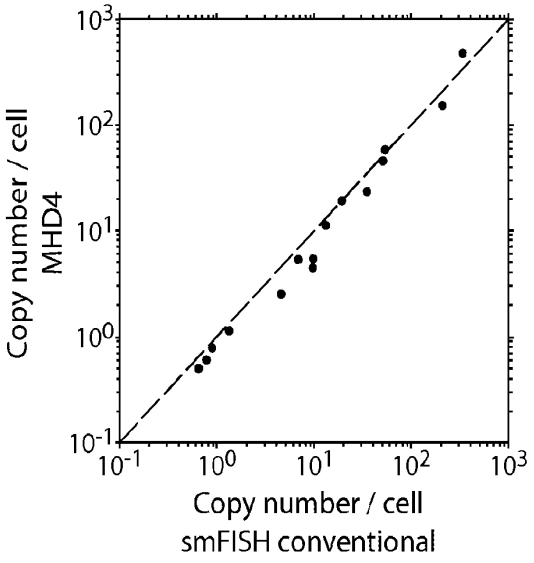


Fig. 18B

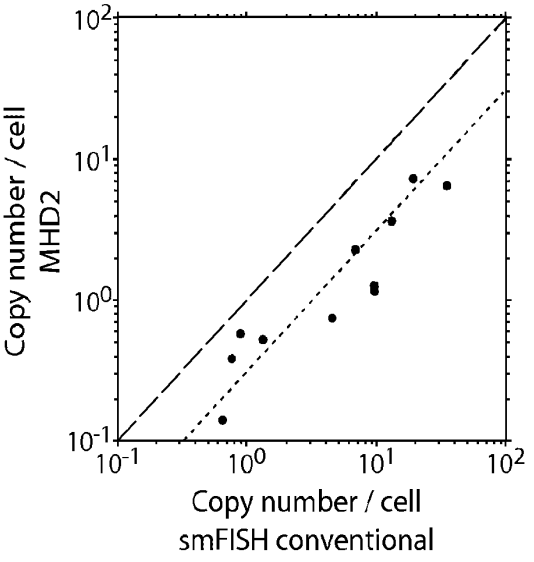


Fig. 18C

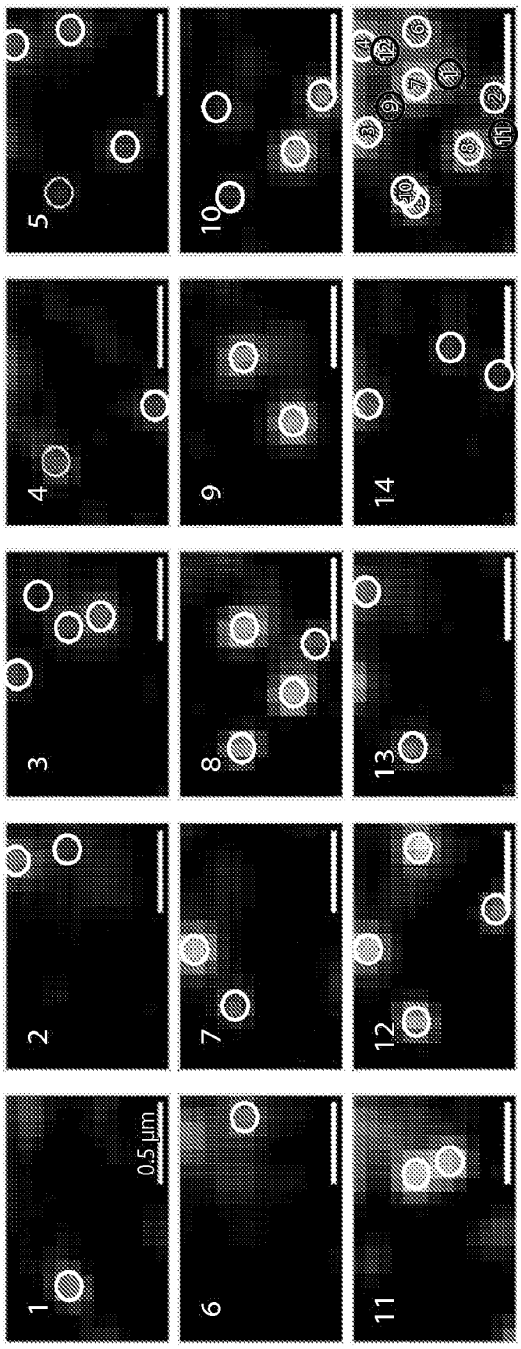


Fig. 19A

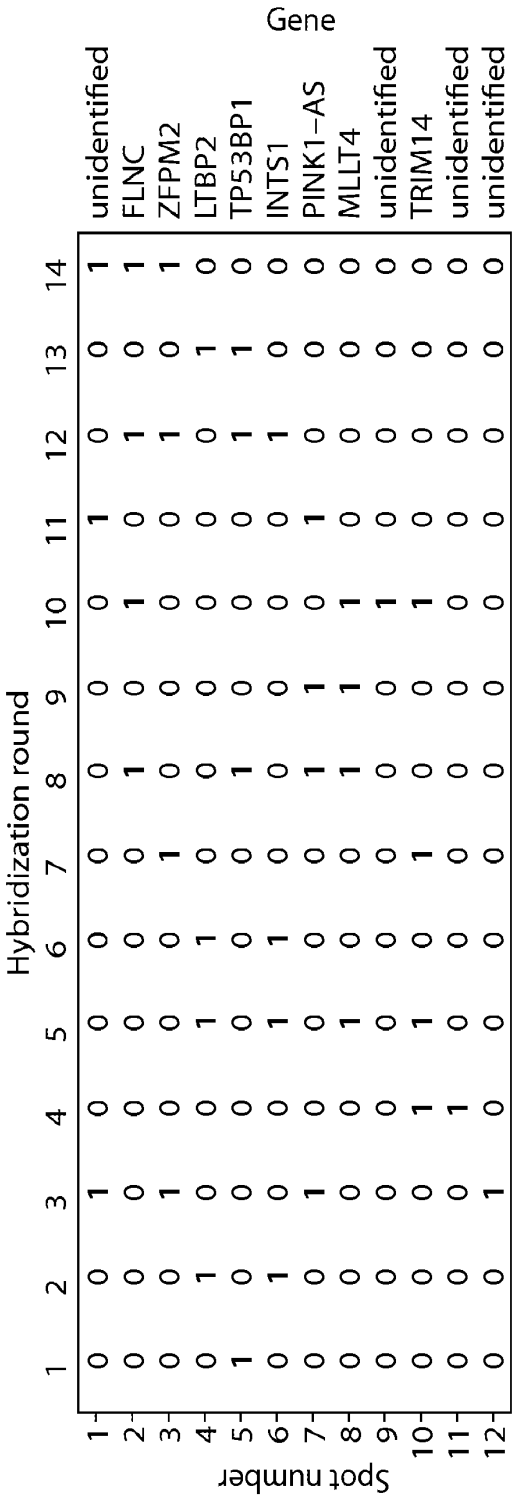

Spot number
Hybridization round
1 2 3 4 5 6 7 8 9 10 11 12 13 14
1 0 0 1 0 0 0 0 0 0 1 0 0 1
2 0 0 0 0 0 0 0 1 0 1 0 1 1
3 0 0 1 0 0 1 0 0 0 0 0 1 1
4 0 1 0 0 1 0 0 0 0 0 0 0 0
5 1 0 0 0 0 0 1 0 0 0 1 1 0
6 0 1 0 0 1 1 0 0 0 0 1 0 0
7 0 0 1 0 0 0 0 1 1 0 0 0 0
8 0 0 0 0 1 0 0 1 1 1 0 0 0
9 0 0 0 0 0 0 0 0 0 1 0 0 0
10 0 0 0 1 1 0 1 0 0 1 0 0 0
11 0 0 0 1 0 0 0 0 0 0 0 0 0
12 0 0 1 0 0 0 0 0 0 0 0 0 0
Gene
unidentified
FLNC
ZFPM2
LTBP2
TP53BP1
INTS1
PINK1-AS
MLLT4
unidentified
TRIM14
unidentified
unidentified

Fig. 19B

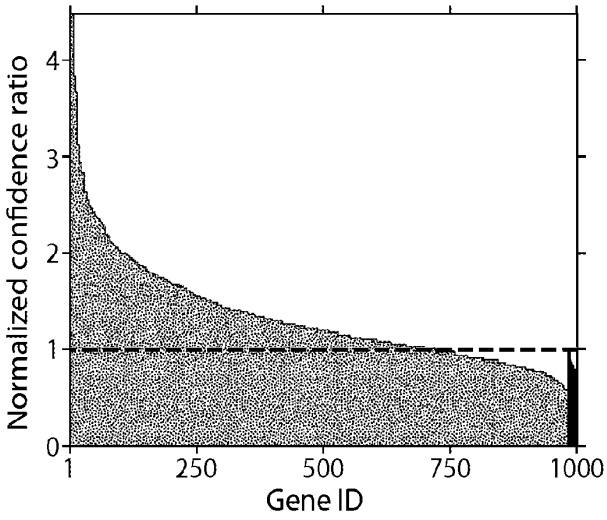


Fig. 19C

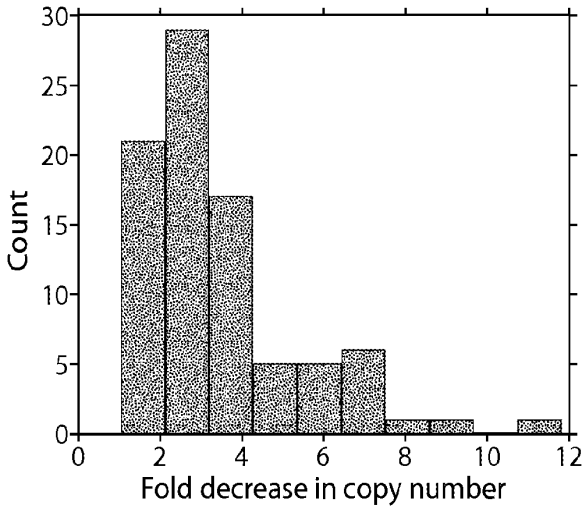


Fig. 19D

Codebook 1 for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
FLNA	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
TNC	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
FBN2	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
KIAA1199	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
COL5A1	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
MALAT1	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
TLN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FBN1	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
SRRM2	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
TEAD1	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
IQGAP1	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
MYH10	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
PAPPA	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
NR4A1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
LRP1	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
SLC38A1	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
SMARCA5	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
CKAP5	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
PIEZO1	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
VCAN	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
PRRC2B	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
SON	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
SLC7A11	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
GPR107	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
KIAA1462	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
notarget001	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
blank001	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0
ROCK2	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
SPTBN1	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FLNC	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
PRPF8	0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
IGF2R	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
FOSB	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
UBR5	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
SVEP1	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
CBX5	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
PDS5A	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
NUMA1	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
PLXNA1	0 0 0 1 0 0 0 0 0 0 0 1 1 0 1 0

Fig. 20-1

Codebook 1 for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
AFAP1	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
SPTAN1	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
DOCK7	0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0
NOTCH2	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
DYNC1H1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
ITGA2	1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0
PRKDC	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
GTF3C1	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
THSD4	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
PTPN14	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
LUZP1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
AMOTL1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
MAN1A2	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
notarget002	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
blank002	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
USP9X	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
CREBBP	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
KPNA4	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
EGFR	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
ZC3H13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
TNRC6A	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
TPR	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
COL7A1	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
ASCC3	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
USP34	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
CENPF	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
KIDINS220	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
SCUBE3	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
NRIP1	0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
NUP98	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
ZFC3H1	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
FAM208B	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
RAB3B	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
AFF4	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
PRKCA	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
FLNB	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
CHD8	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ANKRD52	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
MKI67	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0

Fig. 20-2

Codebook 1 for the 140-gene experiments

Genes	Codewords
DIP2C	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
SSH1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
notarget003	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
blank003	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
DNAJC13	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
DSEL	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PIK3CA	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
MED14	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
HERC2	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
H6PD	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
C14orf132	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
SIPA1L3	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
C17orf51	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
NF1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
VPS13A	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
ABCA2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0
PCNX	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
NKTR	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
FYCO1	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
ATP11A	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
REV3L	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
CASC5	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
AKAP11	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0
CEP250	0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1
USP24	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
CBL	1 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0
RNF169	0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1
TMOD2	0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 0
FASN	0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
notarget004	1 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0
blank004	0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 0 0
UBR2	0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 1
PHIP	1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
GTF3C4	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
QSER1	1 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0
VPS13D	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
FRY	0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 0 0
DIP2B	0 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0
HIVEP2	1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0

Fig. 20-3

Codebook 1 for the 140-gene experiments

Genes	Codewords
DOPEY1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
ALMS1	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
HEATR5B	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
HIVEP1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ARL10	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
HELZ2	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
POLQ	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
CACNA1A	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
KIAA2018	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
LYST	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
ZKSCAN2	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
DIEXF	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
FZD4	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
SLC5A3	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
KIAA1147	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0
LMTK2	0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0
STARD9	0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0
BRCA2	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
notarget005	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20-4

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
FLNA	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
TNC	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
FBN2	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
KIAA1199	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
COL5A1	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
MALAT1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
TLN1	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FBN1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
SRRM2	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
TEAD1	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
IQGAP1	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
MYH10	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
PAPPA	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
NR4A1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
LRP1	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
SLC38A1	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
SMARCA5	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
CKAP5	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
PIEZO1	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
VCAN	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0
PRRC2B	1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
SON	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
SLC7A11	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
GPR107	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
KIAA1462	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
notarget001	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
blank001	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
ROCK2	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
SPTBN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FLNC	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
PRPF8	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
IGF2R	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FOSB	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
UBR5	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
SVEP1	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
CBX5	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
PDS5A	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
NUMA1	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
PLXNA1	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0

Fig. 20-5

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0
AFAP1	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
SPTAN1	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
DOCK7	0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0
NOTCH2	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
DYNC1H1	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
ITGA2	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
PRKDC	0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
GTF3C1	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
THSD4	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
PTPN14	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
LUZP1	1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0
AMOTL1	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
MAN1A2	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
notarget002	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
blank002	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
USP9X	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
CREBBP	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
KPNA4	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
EGFR	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ZC3H13	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
TNRC6A	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
TPR	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
COL7A1	0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 1
ASCC3	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
USP34	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
CENPF	0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 0 0
KIDINS220	0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1
SCUBE3	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
NRIP1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
NUP98	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
ZFC3H1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
FAM208B	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
RAB3B	0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0
AFF4	0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1
PRKCA	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
FLNB	0 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0
CHD8	0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
ANKRD52	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
MKI67	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0

Fig. 20-6

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DIP2C	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
SSH1	0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0
notarget003	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
blank003	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
DNAJC13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
DSEL	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
PIK3CA	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
MED14	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
HERC2	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
H6PD	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
C14orf132	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
SIPA1L3	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
C17orf51	0 1 0 0 0 0 0 0 0 1 1 0 1 0 0 0
NF1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
VPS13A	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
ABCA2	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
PCNX	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
NKTR	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
FYCO1	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
ATP11A	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
REV3L	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
CASC5	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
AKAP11	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
CEP250	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
USP24	1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
CBL	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
RNF169	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
TMOD2	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
FASN	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
notarget004	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
blank004	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
UBR2	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PHIP	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
GTF3C4	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
QSER1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
VPS13D	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
FRY	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
DIP2B	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
HIVEP2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0

Fig. 20-7

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DOPEY1	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
ALMS1	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
HEATR5B	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
HIVEP1	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
ARL10	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
HELZ2	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
POLQ	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
CACNA1A	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
KIAA2018	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
LYST	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
ZKSCAN2	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
DIEXF	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
FZD4	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
SLC5A3	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
KIAA1147	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
LMTK2	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
STARD9	0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0
BRCA2	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
notarget005	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20-8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 (text search)
USPC: 506/9, 16; 536/24.33; 435/6.11 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Patents; Google Scholar
Search terms: nucleic acid probe, hybridization, barcode, index, next generation sequencing, read sequence, target sequence, error correction code, hamming code, fluorescence microscopy, transcriptsome

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/0031243 A1 (CAI et al.) 30 January 2014 (30.01.2014). Especially para [0007], [0018], [0052], [0069], [0141], [0178]	1-3, 88-90,166-171
Y	US 2010/0323348 A1 (HAMADY et al.) 23 January 2010 (23.01.2010).Especially para [0016], [0098].	1-3, 90, 166-171
Y	US 2010/0291557 A1 (LIVAK et al.) 18 November 2010 (18.11.2010). Especially para [0062], [0063]	88-90, 166-171
Y	US 2004/0053300 A1 (SODERLUND et al.) 18 March 2004 (18.03.2004). Especially para [0033], [0035], [0049], [0053]	166-171
X,P	CHEN et al. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science ePub 9 April 2015 Vol 348 No 6233 Pages 1-14. entire article.	1-3, 88-90,166-171

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 November 2015 (16.11.2015)	18 DEC 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-87,99-121,125-135,142,143,148-165,172-176
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
----Go to Extra Sheet for Continuation-----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Claims 1-3, 88-90, 166-171

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

----Continuation of Box III (Lack of Unity of Invention)------

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.

4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]: "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").
-----continued on next sheet-----

---continued from previous sheet---

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe -specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

(PCT Article 17(3)(a) and Rules 40.1 and 40.2(e))

Date of mailing *(day/month/year)* 06 OCT 2015	
Applicant's or agent's file reference H0498.70509	**PAYMENT DUE** within **ONE MONTH** from the above date of mailing
International application No. PCT/US 15/42556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. This International Searching Authority

(i) considers that there are 4 *(number of)* inventions claimed in the international application covered by the claims indicated below/on an extra sheet:
----Go to Extra Sheet for continuation-----

(ii) therefore considers that **the international application does not comply with the requirement of unity of invention** (Rules 13.1, 13.2 and 13.3) for the reasons indicated below/on an extra sheet:
----Go to Extra Sheet for continuation-----

(iii) ☐ has carried out a partial international search (see Annex) ☒ will establish the international search report on those parts of the international application which relate to the invention first mentioned in claims Nos.:
Claims 1-3, 88-90, 166-171

(iv) will establish the international search report on the other parts of the international application only if, and to the extent to which, additional fees are paid.

2. Consequently, the applicant is hereby **invited to pay**, within the time limit indicated above, additional fees in the amount indicated below:

$2,080.00	x	3	=	$6,240.00 (See Item 2 of annex)
Fee per additional invention		number of additional inventions		total amount of additional fees/currency

3. The applicant is informed that, according to Rule 40.2(c), **the payment of any additional fees may be made under protest**, that is, a reasoned statement to the effect that the international application complies with the requirement of unity of invention or that the amount of the required additional fees is excessive, where applicable, subject to the payment of a protest fee.

Where the applicant pays additional fees under protest, the applicant is hereby invited, within the time limit indicated above, to pay a protest fee (Rule 40.2(e)) in the amount of _____ *(amount/currency)*

Where the applicant has not, within the time limit indicated above, paid the required protest fee, the protest will be considered not to have been made and the International Searching Authority will so declare.

4. ☒ Claim(s) Nos. 4-87,99-121,125-135,142,143,148-165,172-176 have been found to be unsearchable under Article 17(2)(b) because of defects under Article 17(2)(a) and therefore have not been included with any invention.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300

Form PCT/ISA/206 (April 2005)

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US 15/42556

Item 2 (continued). For International Applications filed on or after 01 January 2014, Applicant is reminded that the search fee per additional invention indicated in item 2 is the undiscounted fee per additional invention. An Applicant may pay the search fee per additional invention fee reduced by 50% (small entity assertion) or 75% (micro entity certification), as appropriate. See 37 CFR 1.27 and 1.29.

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.

4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore- labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").
-----continued on next sheet-----

Form PCT/ISA/206 (extra sheet) (April 2005)

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US 15/42556

---continued from previous sheet---

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 18 DEC 2015
Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 15/42556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **19 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee W. Young PCT Helpdesk: 571-272-4300 Telephone No. PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2014)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference H0498.70509	FOR FURTHER ACTION	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 15/42556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)	(Earliest) Priority Date *(day/month/year)* 30 July 2014 (30.07.2014)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 9 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

   c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☒ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

   a. the figure of the **drawings** to be published with the abstract is Figure No. 1B

   ☒ as suggested by the applicant.

   ☐ as selected by this Authority, because the applicant failed to suggest a figure.

   ☐ as selected by this Authority, because this figure better characterizes the invention.

   b. ☐ none of the figures is to be published with the abstract.

Form PCT/ISA/210 (first sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-87,99-121,125-135,142,143,148-165,172-176
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
----Go to Extra Sheet for Continuation-----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Claims 1-3, 88-90, 166-171

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 (text search)
USPC: 506/9, 16; 536/24.33; 435/6.11 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Patents; Google Scholar
Search terms: nucleic acid probe, hybridization, barcode, index, next generation sequencing, read sequence, target sequence, error correction code, hamming code, fluorescence microscopy, transcriptsome

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/0031243 A1 (CAI et al.) 30 January 2014 (30.01.2014). Especially para [0007], [0018], [0052], [0069], [0141], [0178]	1-3, 88-90,166-171
Y	US 2010/0323348 A1 (HAMADY et al.) 23 January 2010 (23.01.2010).Especially para [0016], [0098].	1-3, 90, 166-171
Y	US 2010/0291557 A1 (LIVAK et al.) 18 November 2010 (18.11.2010). Especially para [0062], [0063]	88-90, 166-171
Y	US 2004/0053300 A1 (SODERLUND et al.) 18 March 2004 (18.03.2004). Especially para [0033], [0035], [0049], [0053]	166-171
X,P	CHEN et al. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science ePub 9 April 2015 Vol 348 No 6233 Pages 1-14. entire article.	1-3, 88-90,166-171

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 November 2015 (16.11.2015)	18 DEC 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

---Continuation of Box III (Lack of Unity of Invention)-----

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.

4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").
-----continued on next sheet-----

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

----continued from previous sheet---------

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe -specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	18 DEC 2015
Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 15/42556	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule *43bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 *bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	16 November 2015 (16.11.2015)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. --see below---- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

   ☐ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.
4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.
-----continued in Supplemental Box-----

4. Consequently, this opinion has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. Claims 1-3, 88-90, 166-171

Form PCT/ISA/237 (Box No. IV) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	1-3, 88-90, 166-171	NO
Industrial applicability (IA)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO

2. Citations and explanations:

Claims 1-3 lack an inventive step under PCT Article 33(3) as being obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai"), in view of US 2010/0323348 A1 to HAMADY et al. (hereinafter "Hamady").

As to claim 1, Cai teaches exposing a sample to a plurality of nucleic acid probes (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode"), for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample (para [0007]); creating codewords based on the binding of the nucleic acid probes (para [0007]) and for at least some of the codewords, matching the codeword to a valid codeword wherein (para [0007]). Cai does not teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches applying error correction to the codeword to form a valid codeword (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; "Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected"). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine all claim limitations, with the exception of if no match is found, applying error correction to the codeword to form a valid codeword, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled characterizing multiple isoforms of mRNA.

As to claim 2, Cai teaches exposing the sample to at least 5 different nucleic acid probes (para [0052]; "comprises five oligo nucleotide probes").

As to claim 3, Cai further teaches exposing the sample to at least 10 different nucleic acid probes (para [0069]; 8 probes associated with a label emitting a green (G) signal; and 2 probes associated with a label emitting a blue (B) signal").

-----continued in Supplemental Box-----

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
Box IV-3 (Lack of Unity of Invention)

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore- labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V-2 (Citations and Explanations)

Claims 88, 89 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of US 2010/0291557 A1 to LIVAK et al. (hereinafter "Livak").

As to claim 88, Cai teaches method, comprising: exposing a sample to a plurality of nucleic acid probes (para [0007]), and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes (para [0007]) formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences (para [0178]; "Super-resolution barcoding provides a useful single cell follow-up technique to high-throughput sequencing technologies"). Cai does not teach the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample. However, Livak teaches the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences (para [0062-0063]; "the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes specific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay") and for each of the nucleic acid probes determining binding of the target sequences of the nucleic acid and probes within the sample (para [0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Consequently, it would have been obvious for one of ordinary skill in the art to combine exposing a sample to a plurality of nucleic acid probes, at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences, all as taught by Cai, with the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, as taught by Livak, with determining binding of the target sequences of the nucleic acid and probes within the sample, also as taught by Livak, because it would have enabled a method of identifying and quantitating a mRNA population in a cell or group of cells.

As to claim 89, Cai further teaches creating codewords based on determination of the read sequences within the sample (para [0007], [0178]).

Claim 90 lacks an inventive step under PCT Article 33(3) as being obvious over Cai and Livak, in view of Hamady.

As to claim 90, Cai teaches codewords (para [0007]) and matching the codeword to a valid codeword (para [0007]). Neither Cai nor Livak teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches if no match is found, applying error correction to the codeword to form a valid codeword (para [0016], [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine codewords and matching the codeword to a valid codeword, as taught by Cai, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled because it would have enabled characterizing multiple isoforms of mRNA.

--------continued in next Supplemental Box-------

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 and previous Supplemental Box

Claims 166-171 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of Hamady, in further view of Livak, in final view of US 2004/0053300 A1 to SODERLUND et al (hereinafter "Soderlund").

As to claim 166, Cai teaches a method, comprising: associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords (para [0007]) wherein the codewords comprise a number of positions and values for each position (para [0007]). Cai does not teach the codewords form an error-checking and/or error-correcting code space. However, Hamady teaches that the codewords form an error-checking and/or error-correcting code space (para [0016]; [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Cai further teaches forming a plurality of nucleic acid probes (para [0007]) but does not teach that the plurality of probes comprise a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets. However, Livak teaches probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets (para [0062-0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Although Cai further teaches each of the probe pools corresponds to a specific value of a unique position within the codewords (para [0007]; [0162]- In experiments related to those previously described herein, the inventors targeted the YLR414c mRNA with3 sets of oligo probes labeled with different fluorophores"), Cai does not teach grouping the nucleic acid probes into a plurality of probe pools, exposing a sample to one of the probe pools; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions. However, Soderlund teaches grouping the nucleic acid probes into a plurality of probe pools (para [0049]; One or more soluble pools are provided with preset, but optional numbers, preferably varying between 10-500, more preferably between 50-400, most preferably between 100-300 soluble polynucleotide sequences, with distinct sizes allowing their accurate identification, e.g. by mass spectrometry. The soluble probes, which can be identified without any tracer tags, can alternatively be provided with tags"; para [0044]-" Preferred tracer tags are fluorochromes or fluorophors, especially such which have different wave-lengths of emission";. para [0033]-" The term "pool" means a subset or a library of soluble or solubilizable polynucleotide probes. Each pool comprises an optional defined number of polynucleotide probes. A convenient optional number is, for example, approximately 10 probes"; exposing a sample to one of the probe pools (para [0053]; In the basic embodiment of the present invention tracer or primer tagged DNA probes are allowed to hybridize with the preparation of RNA, which is to be analyzed. The analyte polynucleotide sequences or RNA analytes, present in the cell or tissue sample to be determined, are isolated by per se known methods"); Soderlund further teaches repeating the exposing step one or more times, using a different probe pool for at least some of the repetitions (para [0035]; However, because the nucleotide fragments are not characterized, some of the nucleotide fragments obtained by cleaving with various restriction enzymes may by chance have identical sizes. This means that the probes are only partly identifiable by their sizes. When repeating the tests, the redundancy is soon detected and such probes, which cannot be identified with 100% accuracy may be deleted or replaced with other more convenient probes"). Cai further teaches imaging a sample of probes (para [0141]; Imaging of the hybridized cells is carried out on automated fluorescence microscopes"). An artisan of ordinary skill in the art would have recognized that using pools or sets of probes would have enabled obtaining large amounts of information about a given sample, particularly if different probe pools were used in repetitive analysis, as taught by Soderlund. Consequently, it would have been obvious for one of ordinary skill in the art to combine associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, as taught by Cai, with the codewords form an error-checking and/or error-correcting code space, as taught by Hamady, with probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets, as taught by Livak, with grouping the nucleic acid probes into a plurality of probe pools, as taught by Soderlund, with exposing a sample to one of the probe pools; and repeating the exposing one or more times, using a different probe pool for at least some of the repetitions, again as taught by Soderlund, with using imaging technology to visualize the probes, as taught by Cai, because it would have enabled providing an expression profile of a sample.

As to claim 167, Cai further teaches combining the images of the sample into a combined image (para [0165]; "After many iterations of this cycle, a high resolution image can be reconstructed from all of the centroid fits in the individual images").

As to claim 168, Cai further teaches identifying foreground features corresponding to putative nucleic acid targets in the sample (para [0018]; enabling robust hybridization and identification by unique combinations of fluorophores").

As to claim 169, Cai further teaches determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample (para [0007], [0018], [0165]).

As to claim 170, Cai further teaches matching the determined codeword to a valid codeword wherein (para [0007]). Additionally, Hamady further teaches if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target (para [0016], [0098]).

As to claim 171, Cai further teaches imaging the sample using fluorescence microscopy (para [0141]; "Imaging of the hybridized cells is carried out on automated fluorescence microscopes").

Claims 1-3, 88-90, 166-171 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING SUBMISSION, OBTENTION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ÉTATS-UNIS D'AMÉRIQUE

Date of mailing *(day/month/year)* 07 September 2015 (07.09.2015)	
Applicant's or agent's file reference H0498.70509	**IMPORTANT NOTIFICATION**
International application No. PCT/US2015/042556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)
International publication date *(day/month/year)* **Not yet published**	Priority date *(day/month/year)* 30 July 2014 (30.07.2014)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

The applicant is hereby notified of the date of receipt (or of obtaining by the International Bureau) of the priority document(s) relating to all earlier application(s) whose priority is claimed. Unless otherwise indicated by the letters "NR", in the right-hand column or by an asterisk appearing next to the date of receipt, **the priority document concerned was submitted or transmitted to or obtained by the International Bureau in compliance with Rule 17.1(a), (b) or (b-*bis*)**. This Form replaces any previously issued notification concerning submission, transmittal or obtaining of priority documents.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
15 September 2014 (15.09.2014)	62/050,636	US	03 August 2015 (03.08.2015)
30 July 2014 (30.07.2014)	62/031,062	US	03 August 2015 (03.08.2015)
03 April 2015 (03.04.2015)	62/142,653	US	03 August 2015 (03.08.2015)

The letters "NR" denote a priority document which, on the date of mailing of this Form, had not yet been received or obtained by the International Bureau in compliance with Rule 17.1(a), (b) or (b-*bis*). Where the applicant has failed to either submit, request to prepare and transmit, or to request the International Bureau to obtain the priority document within the applicable time limit under that Rule, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

An asterisk "*" next to a date of receipt, denotes a priority document submitted or transmitted to or obtained by the International Bureau but not in compliance with Rule 17.1(a), (b) or (b-*bis*) (the priority document was received after the time limit prescribed in Rule 17.1(a); the request to prepare and transmit the priority document was submitted to the receiving Office after the applicable time limit under Rule 17.1(b) or the request to the International Bureau to obtain the priority document was made after the applicable time limit under Rule 17.1(b-*bis*)). Even though the priority document was not furnished in compliance with Rule 17.1(a), (b) or (b-*bis*), the International Bureau will nevertheless transmit a copy of the document to the designated Offices, for their consideration. In case such a copy is not accepted by the designated Office as the priority document, Rule 17.1(c) provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 70 10

Authorized officer
Mineko Mohri
e-mail pt08.pct@wipo.int
Telephone No. +41 22 338 74 08

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/Michelle Gorman
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
BASIC NATIONAL STAGE FEE	2631	1	140	140
Natl Stage Search Fee - U.S. was the ISA	2641	1	60	60
NATL STAGE EXAM FEE - ALL OTHER CASES	2633	1	360	360
Pages:				
NATL STAGE APPL SZ FEE PER 50 PGS >100	2681	2	200	400
Claims:				
Miscellaneous-Filing:				
Petition:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**960**

Electronic Acknowledgement Receipt

EFS ID:	28187899
Application Number:	15329683
International Application Number:	PCT/US2015/042556
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Michelle Gorman
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	27-JAN-2017
Filing Date:	
Time Stamp:	12:37:38
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$960
RAM confirmation Number	012717INTEFSW12390000
Deposit Account	232825
Authorized User	Wolf Greenfield
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.492 (National application filing, search, and examination fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal of New Application	H049870509US02-TRN-TC.pdf	46344 323bb87a42b9ce1648e50f0e5baf7f9ad5e4291a	no	3
Warnings:					
Information:					
2	Application Data Sheet	H049870509US02-ADS-TC.pdf	31547 b3838ba7d217972e17c7976cb1cea1e6aa434ca4	no	8
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
3		H049870509US02-PAMN-TC.pdf	21121 2d7a86886fc428f64ca1421aec971d64e1e9237b	yes	3

Multipart Description/PDF files in .zip description

Document Description	Start	End
Preliminary Amendment	1	1
Specification	2	2
Applicant Arguments/Remarks Made in an Amendment	3	3

Warnings:					
Information:					
4		H049870509US02-371APL1-TC.pdf	24014301 b5b357d351edde6318a42c7db87301b9c9a285a6	yes	111

Multipart Description/PDF files in .zip description

Document Description	Start	End
Abstract	1	1

	Specification		2	91	
	Claims		92	111	
Warnings:					
Information:					
5		H049870509US02-371APL2-TC.pdf	7241539 eb9757127c8d3eabdfbf089be819e7214f98781c	yes	52
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Drawings-other than black and white line drawings		1	48	
	Documents submitted with 371 Applications		49	52	
Warnings:					
Information:					
6	Documents submitted with 371 Applications	H049870509US02-INV-TC.pdf	698598 c507321a025b96136810aab112bda8e2a6616f7e	no	3
Warnings:					
Information:					
7	Documents submitted with 371 Applications	H049870509US02-ISR-TC.pdf	5548941 9eb255cc92a6c58d38e7abd4e38c4adc8aad896e	no	16
Warnings:					
Information:					
8	Documents submitted with 371 Applications	H049870509US02-304-TC.pdf	235637 f38095f90bedf041effde3f71acfa4a4c038f0a2	no	1
Warnings:					
Information:					
9	Sequence Listing (Text File)	H049870509US02-SEQLST-TC.txt	9884	no	-
Warnings:					
Information:					

10	Fee Worksheet (SB06)	fee-info.pdf	37232 5fdcefd8fc4f6093d1bd0ea0aa82bdb2231e2cd7	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	37885144	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

721259

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2015

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *62/031,062*
FILING DATE: *July 30, 2014*
RELATED PCT APPLICATION NUMBER: *PCT/US15/42556*

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS *US62/031,062*

Certified by

Michelle K. Lee

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PTO/SB/16 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET – Page 1 of 2

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. _____

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Xiaowei	Zhuang	Lexington, Massachusetts
Kok-Hao	Chen	
Alistair	Boettiger	
Jeffrey R.	Moffitt	Somerville, Massachusetts
Siyuan	Wang	Cambridge, Massachusetts

Additional inventors are being named on the _____ separately numbered sheets attached hereto.

TITLE OF THE INVENTION (500 characters max):

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

[X] The address corresponding to Customer Number: 86110

OR

[ ] Firm or Individual Name

Address

City	State	Zip
Country	Telephone	Email

ENCLOSED APPLICATION PARTS (*check all that apply*)

[X] Application Data Sheet. See 37 CFR 1.76.
[ ] CD(s), Number of CDs _____
[X] Drawing(s) *Number of Sheets* 4
[ ] Other (specify) _____
[X] Specification (e.g. description of the invention) *Number of Pages* 73

Fees Due: Filing Fee of $260 ($130 for small entity) ($65 for micro entity). If the specification and drawings exceed 100 sheets of paper, an application size fee is also due, which is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

METHOD OF PAYMENT OF THE FILING FEE AND APPLICATION SIZE FEE FOR THIS PROVISIONAL APPLICATION FOR PATENT

[X] Applicant asserts small entity status. See 37 CFR 1.27.
[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.
[ ] A check or money order made payable to the *Director of the United States Patent and Trademark Office* is enclosed to cover the filing fee and application size fee (if applicable).
[ ] Payment by credit card. Form PTO-2038 is attached.

TOTAL FEE AMOUNT ($)

[ ] The Director is hereby authorized to charge the filing fee and application size fee (if applicable) or credit any overpayment to Deposit Account Number: _____ .

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

3669162.1

PTO/SB/16 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION COVER SHEET - Page 2 of 2

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[X] No.

[ ] Yes, the invention was made by an agency of the United States Government. The U.S. Government agency name is:

[ ] Yes, the invention was made under a contract with an agency of the U.S. Government. The name of the U.S. Government agency and the Government contract number are:

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

SIGNATURE	/Tani Chen/	DATE	July 30, 2014
TYPED or PRINTED NAME	Tani Chen, Sc.D.	REGISTRATION NO. *(if appropriate)*	52,728
TELEPHONE	617.646.8000	DOCKET NUMBER	H0498.70509US00

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within

the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes; exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary

probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid

probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown

where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention; and

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:
GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGATC
ATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCGGACA
AACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG; and

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary and secondary probes are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target nucleic acid, and enough of the

target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then

fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-high resolution imaging techniques (e.g., resolutions better than the wavelength of visible light) may be used. Examples of super-high resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-high resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read

sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For

example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-high resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be

determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the cell or other sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword," e.g., using only first probe 21 and second probe 22 the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probe 21 while the second position corresponds to the binding of probe 22. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction,

double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may by short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein. Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to

produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least

75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid

sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30,

20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a

population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. One of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; fewer than this number of different probes may also be used in

certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, up to $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. Up to 2^n different probes may be possible for n rounds of probes, but the number of nucleic acid probes that are actually used may be any number less than 2^n. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid probe codewords. If a match is found, then the nucleic acid probe can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus the correct nucleic acid probe. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one of the nucleic acid probes is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one of the nucleic acid probes is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one

set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions.

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated

Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be

determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633

nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. . Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a

stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO

633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low

calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red,

Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding, error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular

(having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled "Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified

from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the

oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective smFISH approach, but also allows researchers to benefit

from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a

different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by

hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. Even for a modest number of moderately expressed mRNAs, the density of labeled mRNAs exceed that at which each mRNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of secondary hybridization and imaging, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from

individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps (meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be

included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging of super-resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended

Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into

two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 °C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10

B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to

synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-

microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters

may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0]. Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are

meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also

including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but

not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

CLAIMS

1. A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2. The method of claim 1, comprising exposing the sample to at least 5 different nucleic acid probes.

3. The method of any one of claims 1 or 2, comprising exposing the sample to at least 10 different nucleic acid probes.

4. The method of any one of claims 1-3, comprising exposing the sample to at least 100 different nucleic acid probes.

5. The method of any one of claims 1-4, comprising exposing the sample to a plurality of nucleic acid probes simultaneously.

6. The method of any one of claims 1-5, comprising exposing the sample to a plurality of nucleic acid probes sequentially.

7. The method of claims 1-6, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8. The method of claim 7, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of RNA species in the sample.

9. The method of claim 7 or 8, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of DNA sequences in the sample.

10. The method of any one of claims 1-9, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11. The method of claim 10, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

12. The method of any one of claims 10 or 11, wherein the target sequence is substantially complementary to a nucleic acid sequence encoding a protein.

13. The method of any one of claims 10-12, wherein the target sequence is substantially complementary to messenger RNA (mRNA).

14. The method of any one of claims 10-13, wherein the plurality of nucleic acid probes comprises at least 8 possible read sequences.

15. The method of any one of claims 10-14, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

16. The method of any one of claims 10-15, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

17. The method of any one of claims 10-16, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

18. The method of any one of claims 10-17, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19. The method of any one of claims 10-18, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

20. The method of any one of claims 10-19, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

21. The method of any one of claims 10-20, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

22. The method of any one of claims 10-21, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

23. The method of any one of claims 10-22, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

24. The method of any one of claims 10-23, wherein at least some of the plurality of nucleic acid probes comprise no more than 3 read sequences.

25. The method of any one of claims 10-24, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

26. The method of claims 1-25, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27. The method of claim 26, wherein the first signaling entity is fluorescent.

28. The method of any one of claims 26 or 27, wherein the first signaling entity is a protein.

29. The method of any one of claims 26-28, wherein the first signaling entity is a dye.

30. The method of any one of claims 26-29, wherein the first signaling entity is a nanoparticle.

31. The method of any one of claims 26-30, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32. The method of claim 31, wherein the first signaling entity and the second signaling entity are identical.

33. The method of claim 31, wherein the first signaling entity and the second signaling entity are not identical.

34. The method of any one of claims 31-33, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35. The method of claim 34, comprising inactivating the first signaling entity by photobleaching at least some of the first signaling entity.

36. The method of any one of claims 34 or 35, comprising inactivating the first signaling entity by chemically bleaching at least some of the first signaling entity.

37. The method of any one of claims 34-36, comprising inactivating the first signaling entity by exposing the first signaling entity to a reactant able to alter the structure of the signaling entity.

38. The method of any one of claims 34-37, comprising inactivating the first signaling entity by removing at least some of the first signaling entity.

39. The method of any one of claims 34-38, comprising inactivating the first signaling entity by dissociating the first signaling entity from the first secondary probe.

40. The method of any one of claims 34-39, comprising inactivating the first signaling entity by dissociating the first secondary probe that contains the first signaling entity from the sample.

41. The method of any one of claims 34-40, comprising inactivating the first signaling entity by exposing the signaling entity or the first secondary probe to an enzyme.

42. The method of claims 1-41, wherein at least some of the plurality of nucleic acid probes comprise DNA.

43. The method of any one of claims 1-42, wherein at least some of the plurality of nucleic acid probes comprise RNA.

44. The method of any one of claims 1-43, wherein at least some of the plurality of nucleic acid probes comprise PNA.

45. The method of any one of claims 1-44, wherein at least some of the plurality of nucleic acid probes comprise LNA.

46. The method of any one of claims 1-45, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

47. The method of any one of claims 1-46, wherein at least some of the plurality of nucleic acids probes are configured to bind to a nucleic acid within the sample.

48. The method of any one of claims 1-47, wherein at least some of the binding of the nucleic acid probes to a target within the sample is specific binding.

49. The method of any one of claims 1-48, wherein at least some of the binding of the nucleic acid probes to a target within the sample is via Watson-Crick base pairing.

50. The method of any one of claims 1-49, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

51. The method of any one of claims 1-50, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

52. The method of any one of claims 1-51, wherein at least some of the plurality of nucleic acid probes are configured to bind to mRNA.

53. The method of any one of claims 1-52, wherein at least some of the plurality of nucleic acid probes are configured to bind to transfer RNA (tRNA).

54. The method of any one of claims 1-53, wherein at least some of the plurality of nucleic acid probes are configured to bind to ribosomal RNA (rRNA).

55. The method of any one of claims 1-54, wherein at least some of the plurality of nucleic acid probes are configured to bind to lncRNA.

56. The method of any one of claims 1-55, wherein at least some of the plurality of nucleic acid probes are configured to bind to snoRNA.

57. The method of any one of claims 1-56, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

58. The method of any one of claims 1-57, wherein the plurality of nucleic acids probes is configured to bind to DNA.

59. The method of any one of claims 1-58, wherein at least some of the plurality of nucleic acid probes are configured to bind to genomic DNA.

60. The method of any one of claims 1-59, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

61. The method of any one of claims 1-60, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm.

62. The method of any one of claims 1-61, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 80 nm.

63. The method of any one of claims 1-62, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 50 nm.

64. The method of any one of claims 1-63, wherein the sample comprises a cell.

65. The method of claim 64, wherein the cell is a human cell.

66. The method of any one of claims 64 or 65, wherein the cell is fixed.

67. The method of any one of claims 1-66, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

68. The method of any one of claims 1-67, comprising determining binding of the nucleic acid probes using an optical imaging technique.

69. The method of any one of claims 1-68, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

70. The method of any one of claims 1-69, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

71. The method of claim 70, comprising determining binding of the nucleic acid probes using stochastic optical reconstruction microscopy (STORM).

72. The method of any one of claims 70 or 71, comprising determining binding of the nucleic acid probes using photoactivated localization microscopy (PALM) or fluorescence photoactivation localization microscopy (FPALM).

73. The method of any one of claims 70-72, comprising determining binding of the nucleic acid probes using stimulated emission depletion microscopy (STED).

74. The method of any one of claims 70-73, comprising determining binding of the nucleic acid probes using structured illumination microscopy (SIM).

75. The method of any one of claims 70-74, comprising determining binding of the nucleic acid probes using reversible saturated optically linear fluorescence transition (RESOLFT) microscopy.

76. A method, comprising:

exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and

for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

77. The method of claim 76, further comprising:

creating codewords based on determination of the read sequences within the sample.

78. The method of claim 77, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

79. A method, comprising:

exposing a sample to a plurality of primary nucleic acid probes;

exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample;

creating codewords based on fluorescence of the secondary nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

80. A method, comprising:

exposing a plurality of primary nucleic acid probes to a sample; and

exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample, wherein at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

81. The method of claim 80, further comprising:

creating codewords based on fluorescence of the secondary nucleic acid probes.

82. The method of claim 81, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

83. A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

84. The composition of claim 83, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

85. The composition of claim 84, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 3.

86. The composition of any one of claims 84 or 85, wherein the code is a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, or a Hamming(127, 120) code.

87. The composition of any one of claims 83-86, wherein the code is a SECDED code.

88. The composition of claim 87, wherein the code is a SECDED(8,4) code.

89. The composition of claim 87, wherein the code is a SECDED(16,4) code.

90. The composition of claim 87, wherein the code is a SECDED(16, 11) code, a SECDED (22, 16) code, a SECDED (39, 32) code, or a SECDED (72, 64) code.

91. The composition of any one of claims 87-90, wherein only codewords with a constant number of ones are used.

92. The composition of any one of claims 83-91, wherein the plurality of nucleic acid probes comprises no more than 100 possible read sequences.

93. The composition of any one of claims 83-92, wherein the plurality of nucleic acid probes comprises no more than 64 possible read sequences.

94. The composition of any one of claims 83-93, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

95. The composition of any one of claims 83-94, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

96. The composition of any one of claims 83-95, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

97. The composition of any one of claims 83-96, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

98. The composition of any one of claims 83-97, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

99. The composition of any one of claims 83-98, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

100. The composition of any one of claims 83-99, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

101. The composition of any one of claims 83-100, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

102. The composition of any one of claims 83-101, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

103. The composition of any one of claims 83-102, wherein at least some of the plurality of nucleic acid probes comprise DNA.

104. The composition of any one of claims 83-103, wherein at least some of the plurality of nucleic acid probes comprise RNA.

105. The composition of any one of claims 83-104, wherein at least some of the plurality of nucleic acid probes comprise PNA.

106. The composition of any one of claims 83-105, wherein at least some of the plurality of nucleic acid probes comprise LNA.

107. The composition of any one of claims 83-106, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

108. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position

within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

109. The method of claim 108, wherein the code space has a Hamming distance of at least 2.

110. The method of any one of claims 108 or 109, wherein the plurality of targets comprise RNA.

111. A method, comprising:

exposing nucleic acid probes formed using the method of any one of claims 108-110 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets within the sample.

112. The method of claim 111, wherein at least some of the plurality of secondary nucleic probes is able to bind to a read sequence of a nucleic acid probe.

113. The method of claim 112, wherein at least some of the plurality of secondary nucleic probes is able to specifically bind to only one read sequence.

114. The method of claim 111-113, wherein if no match is found between a codeword and the codeword derived from the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

115. The method of claim 111-114, wherein at least some of the secondary nucleic acid probes each comprise a signaling entity.

116. The method of claim 115, comprising imaging the signaling entity to determine binding of the plurality of secondary nucleic probes within the sample.

117. The method of claim 116, comprising imaging the signaling entity using a super-resolution imaging technique.

118. The method of any one of claims 116 or 117, further comprising applying drift correction to the images.

119. The method of any one of claims116-118, further comprising applying physical registration to the images.

120. The method of claim 111-119, comprising determining binding of the plurality of secondary nucleic probes within the sample at a resolution better than 100 nm.

121. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

122. The method of claim 121, where in the codewords comprise a number of positions that is less than the number of targets.

123. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a

plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

124. The method of claim 123, wherein the codewords form an error-checking space.

125. The method of any one of claims 123 or 124, wherein the codewords form an error-correcting code space.

126. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

127. The method of any one of claims 123-126, wherein the groups are applied to the sample sequentially using fluidic devices.

128. The method of any one of claims 123-127, wherein each group of nucleic acid probes contains a group identification sequence distinguishable from the group identification sequence of the other groups.

129. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code;

associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample;

forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences;

forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences;

exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample;

exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

130. The method of claim 129, wherein the wherein the read sequences are taken from a set of orthogonal sequences, which have a homology of less than 15 basepairs with one another and with the nucleic acid species in a sample.

131. The method of any one of claims 129 or 130, comprising combining the images of the sample into a combined image.

132. The method of claim 131, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

133. The method of claim 132, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

134. The method of claim 133, comprising matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword that correspond to a nucleic acid target

135. The method of any one of claims 129-134, comprising imaging the sample at a resolution better than 300 nm.

136. The method of any one of claims 129-135, comprising imaging the sample at a resolution better than 100 nm.

137. The method of any one of claims 129-136, comprising imaging the sample using a super-resolution fluorescence imaging technique.

138. The method of any one of claims 129-137, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

139. The method of any one of claims 129-138, where in the codewords comprise a number of positions that is less than the number of targets.

140. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes comprising a signaling entity and a

target sequence that uniquely binds to one of the nucleic acid targets;

grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords;

exposing a sample to one of the probe pools;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

141. The method of claim 140, comprising combining the images of the sample into a combined image.

142. The method of claim 141, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

143. The method of claim 142, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

144. The method of claim 143, comprising matching the determined codeword to a valid codeword wherein, if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target.

145. The method of any one of claims 140-144, comprising imaging the sample at a resolution better than 300 nm.

146. The method of any one of claims 140-145, comprising imaging the sample at a resolution better than 100 nm.

147. The method of any one of claims 140-146, comprising imaging the sample using a super-resolution fluorescence imaging technique.

148. The method of any one of claims 140-147, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

5

149. The method of any one of claims 140-148, wherein the codewords comprise a number of positions that is less than the number of targets.

ABSTRACT

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Xiaowei
Family Name::	Zhuang
City of Residence::	Lexington
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	5 Luongo Farm Lane
City of mailing address::	Lexington
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02421

Inventor Number::	2
Given Name::	Kok-Hao
Family Name::	Chen

Inventor Number::	3
Given Name::	Alistair
Family Name::	Boettiger

Inventor Number::	4
Given Name::	Jeffrey

Middle Name::	R.
Family Name::	Moffitt
City of Residence::	Somerville
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	175 Beacon Street
	Apt. #108
City of mailing address::	Somerville
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02143

Inventor Number::	5
Given Name::	Siyuan
Family Name::	Wang
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	28 Fernald Drive
	Apt. #31
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02138

Correspondence Information

Correspondence Customer Number::	86110

Application Information

Application Type::	Provisional
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No
Title::	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Attorney Docket Number::	H0498.70509US00
Request for Early Publication?::	No
Request for Non-Publication?::	No
Small Entity?::	Yes
Petition included?::	No
Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2::	No
Authorization to Permit Access to the Instant Application by the Participating Offices::	No

Representative Information

Representative Customer Number::	86110

Domestic Priority Information

Foreign Priority Information

Applicant Information

Assignee Number::	1
Applicant Type::	Assignee – obligated to assign to
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02138-5701

Assignee Information including Non-Applicant Assignee Information

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.			
Signature	/Tani Chen/	Date	July 30, 2014
Name (Print/Type)	Tani Chen, Sc.D.	Registration No. (Attorney/Agent)	52,728

Encoding Scheme

RNA target	codeword	probe sequence
mRNA I	110	
mRNA II	011	
mRNA III	101	

Primary sites (RNA tag)
Secondary site 1
Secondary site 2
Secondary site 3

Fig. 1A

b) **Code-Reading Scheme 1: Primaries and Secondaries**

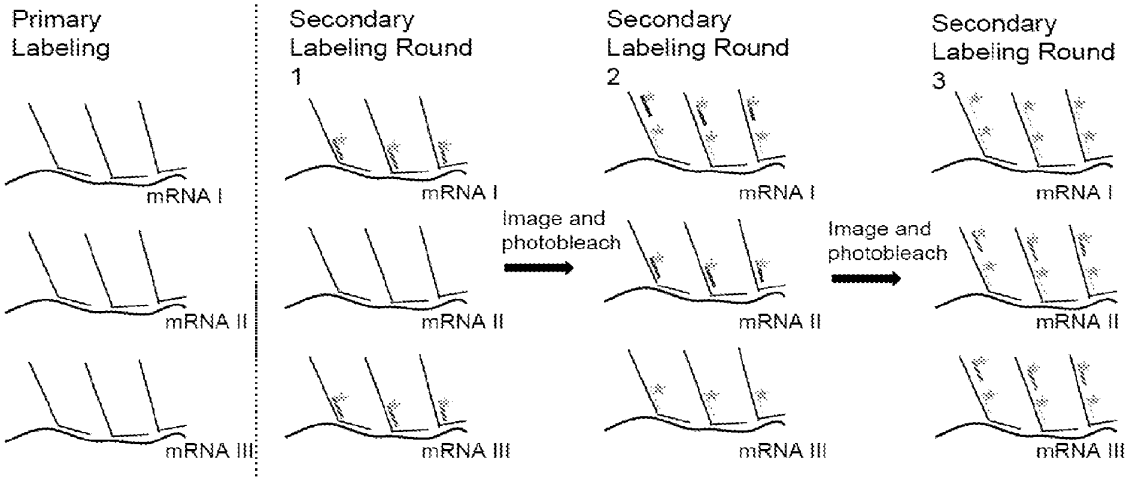


Fig. 1B

Code-Reading Scheme 2: Primaries Only

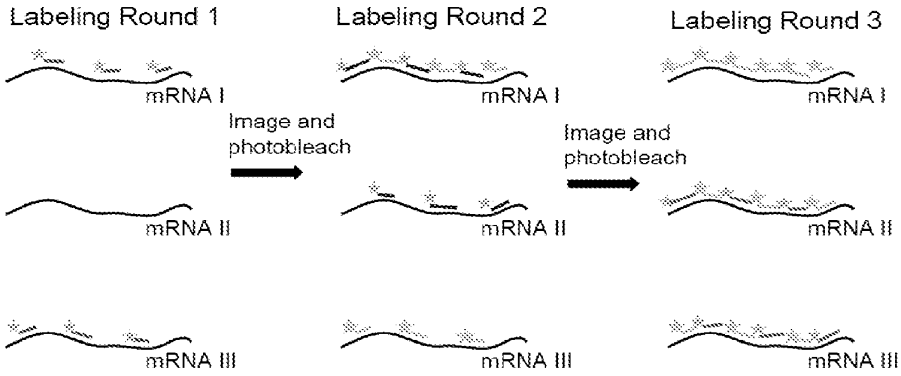


Fig. 1C

Fig. 2A

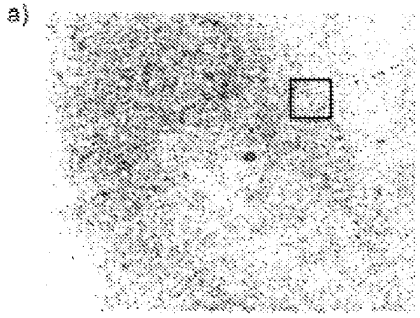


Fig. 2B

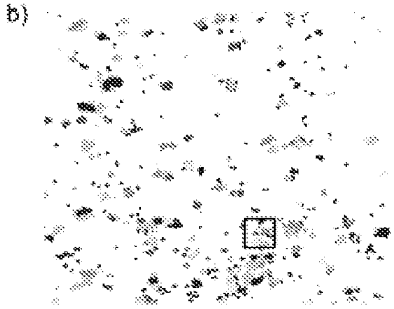
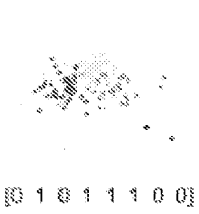


Fig. 2C



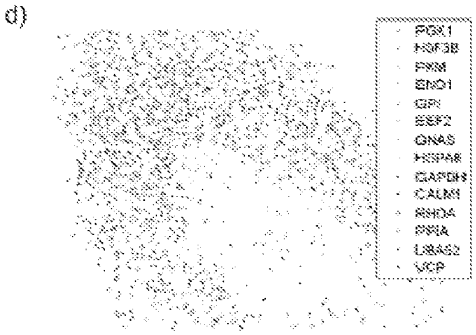


Fig.2D

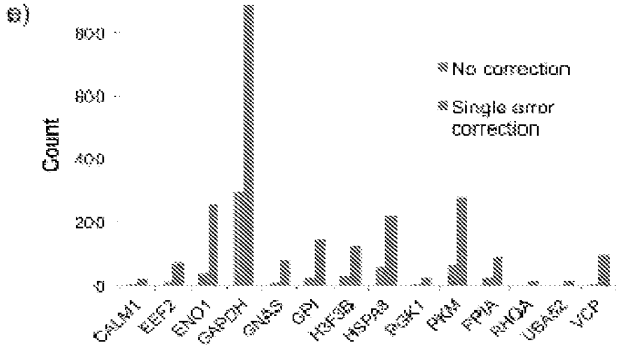


Fig. 2E

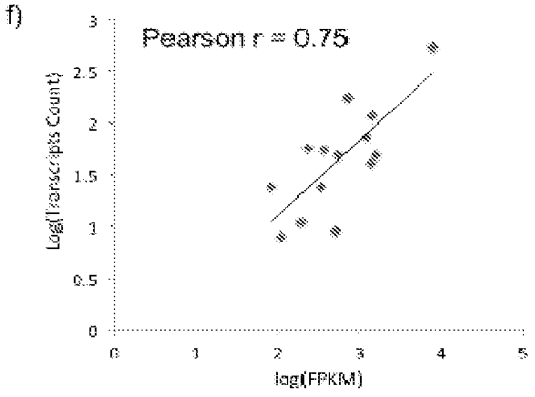


Fig. 2F

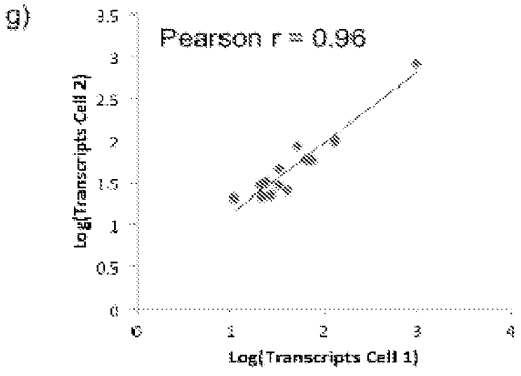


Fig. 2G

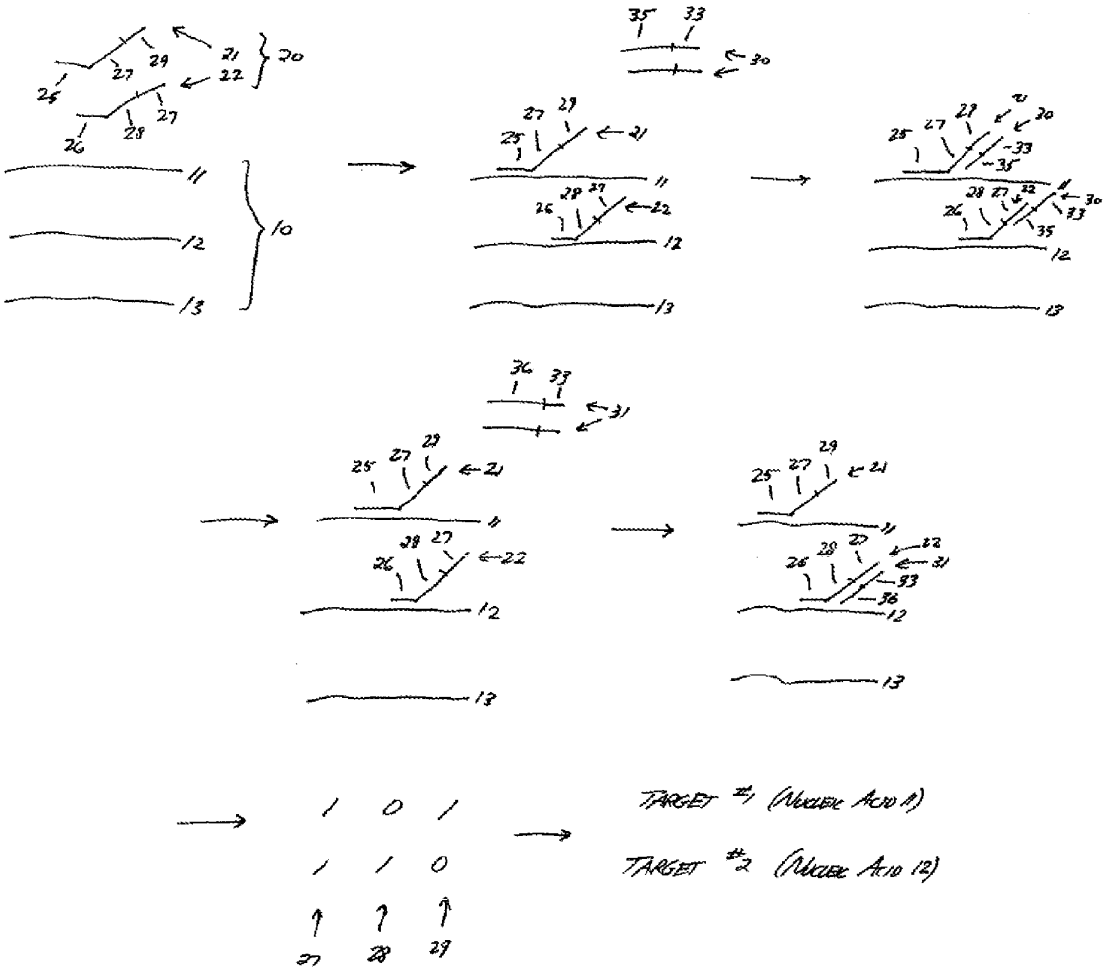


Fig. 3A

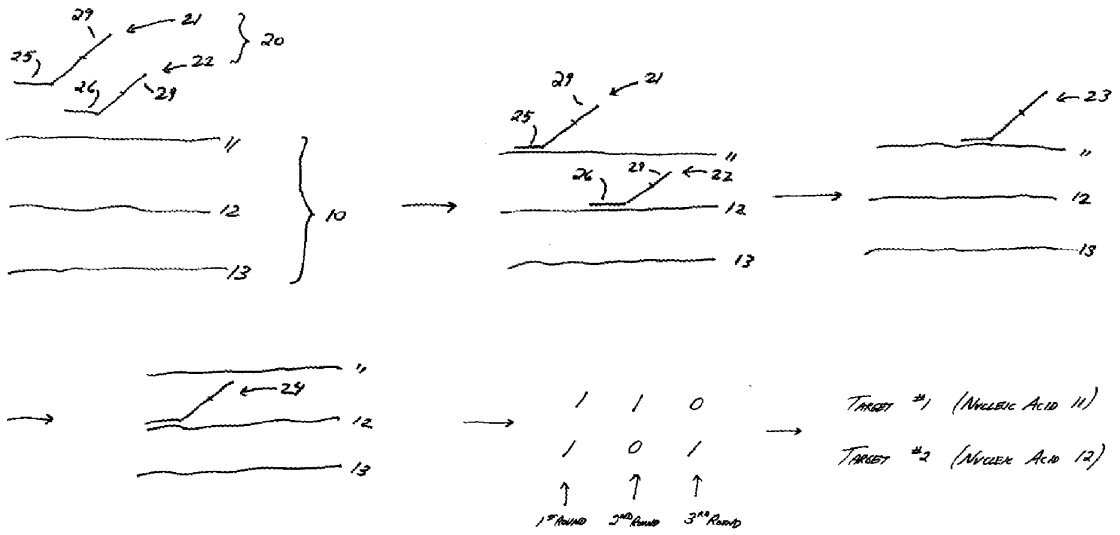


Fig. 3B

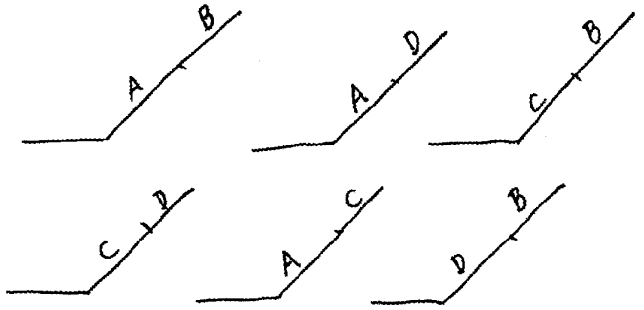


Fig. 4A

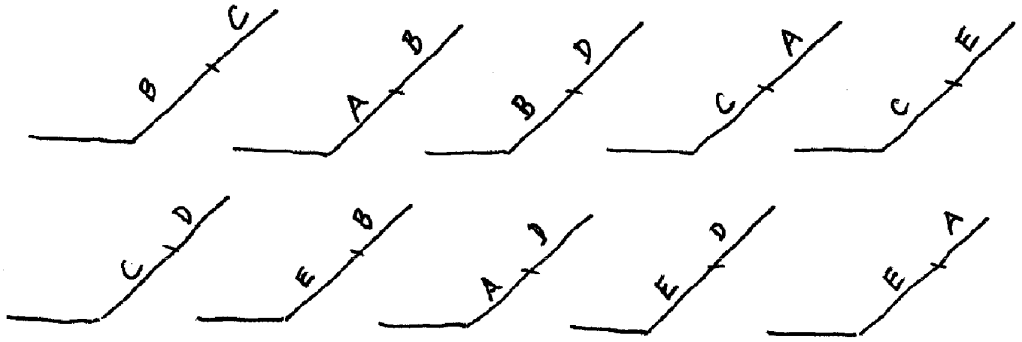


Fig. 4B

Electronic Acknowledgement Receipt

EFS ID:	19733759
Application Number:	62031062
International Application Number:	
Confirmation Number:	3899
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Trish McDonald
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US00
Receipt Date:	30-JUL-2014
Filing Date:	
Time Stamp:	18:54:44
Application Type:	Provisional

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Provisional Cover Sheet (SB16)	H049870509US00-PROCVR-TC.pdf	38684 4f5fa6f63a5bb7f92ee86164d7542598a0eefb9d	no	2

Warnings:

This is not a USPTO supplied Provisional Cover Sheet SB16 form.

Information:					
2	Application Data Sheet	H049870509US00-ADS-TC.pdf	52374 ece51a97cb3e24f0fb062738b0e6a97f14599435	no	4
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
3		H049870509US00-PROAPL-TC.pdf	228738 0efb49b23d45d61c4a5c0b9592700fd5eb69f776	yes	73

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Specification	1	53
Claims	54	72
Abstract	73	73

Warnings:					
Information:					
4	Drawings-other than black and white line drawings	H049870509US00-DRW-TC.pdf	511020 8df0daea93933458680bd9f4119bb84a45af1317	no	4
Warnings:					
Information:					
		Total Files Size (in bytes):	830816		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 62031062 Document Date: 07/30/2014

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

DOCUMENT MADE AVAILABLE UNDER THE PATENT COOPERATION TREATY (PCT)

International application number:		**PCT/US2015/042556**
International filing date:		**29 July 2015 (29.07.2015)**
Document type:		**Certified copy of priority document**
Document details:	Country/Office:	**US**
	Number:	**62/031,062**
	Filing date:	**30 July 2014 (30.07.2014)**
Date of receipt at the International Bureau:		**03 August 2015 (03.08.2015)**

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a),(b) or (b-*bis*)

34, chemin des Colombettes
1211 Geneva 20, Switzerland
www.wipo.int

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

PCT/US15/42556
International Application No.

29 JUL 2015 (29.07.15)
International Filing Date

RO/US
Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* H0498.70509

Box No. I TITLE OF INVENTION

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

Box No. II APPLICANT ☐ This person is also inventor

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

PRESIDENT AND FELLOWS OF HARVARD COLLEGE
17 Quincy Street
Cambridge, Massachusetts 02138
United States of America

Telephone No.

Facsimile No.

Applicant's registration No. with the Office

E-mail authorization: Marking one of the check-boxes below authorizes the receiving Office, the International Searching Authority, the International Bureau and the International Preliminary Examining Authority to use the e-mail address indicated in this Box to send, notifications issued in respect of this international application to that e-mail address if those offices are willing to do so.

☐ as advance copies followed by paper notifications; or ☐ exclusively in electronic form (no paper notifications will be sent).

E-mail address:

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☒ all designated States ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☒ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

Telephone No. 617.646.8000

Facsimile No. 617.646.8646

Agent's registration No. with the Office 52,728

E-mail authorization: Marking one of the check-boxes below authorizes the receiving Office, the International Searching Authority, the International Bureau and the International Preliminary Examining Authority to use the e-mail address indicated in this Box to send, notifications issued in respect of this international application to that e-mail address if those offices are willing to do so.

☒ as advance copies followed by paper notifications; or ☐ exclusively in electronic form (no paper notifications will be sent).

E-mail address: PCT@wolfgreenfield.com

☐ **Address for correspondence:** Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

ZHUANG, Xiaowei
5 Luongo Farm Lane
Lexington, Massachusetts 02421
United States of America

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [x] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

CHEN, Kok-Hao
37 Jalan Lepas
Singapore 537961
Republic of Singapore

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [x] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

BOETTIGER, Alistair
1550 Cambridge Street
Apt. #2
Cambridge, Massachusetts 02139
United States of America

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [x] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

MOFFITT, Jeffrey R.
175 Beacon Street
Apt. #108
Somerville, Massachusetts 02143
United States of America

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [x] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

- [x] Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

WANG, Siyuan
28 Fernald Drive
Apt. #31
Cambridge, Massachusetts 02138
United States of America

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [x] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [ ] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [ ] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- [ ] applicant only
- [ ] applicant and inventor
- [ ] inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of:
- [ ] all designated States
- [ ] the States indicated in the Supplemental Box

- [ ] Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No. V DESIGNATIONS

The filing of this request **constitutes under Rule 4.9(a) the designation** of all Contracting States bound by the PCT on the international filing date, for the grant of every kind of protection available and, where applicable, for the grant of both regional and national patents.

However,

☐ DE Germany **is not designated** for any kind of national protection

☐ JP Japan **is not designated** for any kind of national protection

☐ KR Republic of Korea **is not designated** for any kind of national protection

(The check-boxes above may only be used to exclude (irrevocably) the designations concerned if, at the time of filing or subsequently under Rule 26bis.1, the international application contains in Box No. VI a priority claim to an earlier national application filed in the particular State concerned, in order to avoid the ceasing of the effect, under the national law, of this earlier national application.)

Box No. VI PRIORITY CLAIM AND DOCUMENT

The priority of the following earlier application(s) is hereby claimed:

Filing date of earlier application *(day/month/year)*	Number of earlier application	Where earlier application is:		
		national application: country or Member of WTO	regional application: regional Office	international application: receiving Office
item (1) 30 July 2014 (30.07.2014)	62/031062	US		
item (2) 15 September 2014 (15.09.2014)	62/050636	US		
item (3) 03 April 2015 (03.04.2015)	62/142653	US		

☐ Further priority claims are indicated in the Supplemental Box.

Furnishing the priority document(s):

☒ The **receiving Office** is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application(s) was filed with the receiving Office which, for the purposes of this international application, is the receiving Office)* identified above as:

☒ all items ☒ item (1) ☒ item (2) ☒ item (3) ☐ other, see Supplemental Box

☐ The **International Bureau** is requested to obtain from a digital library a certified copy of the earlier application(s) identified above, using, where applicable, the access codes(s) indicated below *(if the earlier application(s) is available to it from a digital library)*:

☐ item (1) access code _____ ☐ item (2) access code _____ ☐ item (3) access code _____ ☐ other, see Supplemental Box

Restore the right of priority: the receiving Office is requested to restore the right of priority for the earlier application(s) identified above or in the Supplemental Box as item(s) (_____). *(See also the Notes to Box No. VI; further information **must** be provided to support a request to restore the right of priority.)*

Incorporation by reference: where an element of the international application referred to in Article 11(1)(iii)(d) or (e) or a part of the description, claims or drawings referred to in Rule 20.5(a) is not otherwise contained in this international application but is completely contained in an earlier application whose priority is claimed on the date on which one or more elements referred to in Article 11(1)(iii) were first received by the receiving Office, that element or part is, subject to confirmation under Rule 20.6, incorporated by reference in this international application for the purposes of Rule 20.6.

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) *(if more than one International Searching Authority is competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used)*:

ISA / US

Box No. IX CHECK LIST for EFS-Web filings - this sheet is only to be used when filing an international application with RO/US via **EFS-Web**

This international application **contains** the following:	Number of sheets
(a) request form PCT/RO/101 (including any declarations and supplemental sheets)	5
(b) description (excluding any sequence listing part of the description, see (f), below)	90
(c) claims	20
(d) abstract	1
(e) drawings (if any)	17
(f) sequence listing part of the description in the form of an **image file** (e.g. PDF)	
Total number of sheets (including the sequence listing part of the description if **filed as an image file**)	133

(g) sequence listing part of the description

[x] filed in the form of an **Annex C/ST.25 text file**

[ ] WILL BE filed separately on physical data carrier(s), on the same day and in the form of an **Annex C/ST.25 text file**

Indicate type and number of physical data carrier(s)

This international application is **accompanied by** the following item(s) *(mark the applicable check-boxes below and indicate in right column the number of each item):*	Number of items
1. [x] fee calculation sheet	1
2. [ ] original separate power of attorney	
3. [ ] original general power of attorney	
4. [ ] copy of general power of attorney; reference number:	
5. [ ] priority document(s) identified in Box No. VI as item(s)	
6. [ ] Translation of international application into *(language)*:	
7. [ ] separate indications concerning deposited microorganism or other biological material	
8. [ ] *(only where item (f) is marked in the left column)* copy of the sequence listing in electronic form (Annex C/ST.25 text file) not forming part of the international application but **furnished only for the purposes of international search** under Rule 13*ter*	
9. [ ] *(only where item (f) (in the left column) and item 8 (above) are marked)* a statement confirming that "the information recorded in electronic form submitted under Rule 13*ter* is identical to the sequence listing as contained in the international application" as filed via EFS-Web:	
10. [ ] copy of results of earlier search(es) (Rule 12*bis*.1(a))	
11. [x] other *(specify)*: Transmittal Letter	1

Figure of the drawings which should accompany the abstract: 1b

Language of filing of the international application: English

Box No. X SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

Attorney Reg. No. (when applicable): 52,728

/Tani Chen/
Tani Chen, Sc.D., Agent for Applicant(s)
WOLF, GREENFIELD & SACKS, P.C.

For receiving Office use only

1. Date of actual receipt of the purported international application:	29-07-2015	2. Drawings:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:		[ ] received:
4. Date of timely receipt of the required corrections under PCT Article 11(2):		[ ] not received:
5. International Searching Authority (if two or more are competent): ISA / US	6. [ ] Transmittal of search copy delayed until search fee is paid	

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2015/042556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)	Priority date *(day/month/year)* 30 July 2014 (30.07.2014)
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 10 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☒	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☒	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☒	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 31 January 2017 (31.01.2017)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Mineko Mohri e-mail: pct.team8@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	18 DEC 2015
Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 15/42556	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule *43bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 *bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	16 November 2015 (16.11.2015)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. [x] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:
   - a. [x] forming part of the international application as filed:
     - [x] in the form of an Annex C/ST.25 text file.
     - [ ] on paper or in the form of an image file.
   - b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   - c. [ ] furnished subsequent to the international filing date for the purposes of international search only:
     - [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
     - [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Form PCT/ISA/237 (Box No. I) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. --see below---- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

   ☐ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.
4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.
-----continued in Supplemental Box-----

4. Consequently, this opinion has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. Claims 1-3, 88-90, 166-171

Form PCT/ISA/237 (Box No. IV) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	1-3, 88-90, 166-171	NO
Industrial applicability (IA)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO

2. Citations and explanations:

Claims 1-3 lack an inventive step under PCT Article 33(3) as being obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai"), in view of US 2010/0323348 A1 to HAMADY et al. (hereinafter "Hamady").

As to claim 1, Cai teaches exposing a sample to a plurality of nucleic acid probes (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode"), for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample (para [0007]); creating codewords based on the binding of the nucleic acid probes (para [0007]) and for at least some of the codewords, matching the codeword to a valid codeword wherein (para [0007]). Cai does not teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches applying error correction to the codeword to form a valid codeword (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; "Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected"). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine all claim limitations, with the exception of if no match is found, applying error correction to the codeword to form a valid codeword, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled characterizing multiple isoforms of mRNA.

As to claim 2, Cai teaches exposing the sample to at least 5 different nucleic acid probes (para [0052]; "comprises five oligo nucleotide probes").

As to claim 3, Cai further teaches exposing the sample to at least 10 different nucleic acid probes (para [0069]; 8 probes associated with a label emitting a green (G) signal; and 2 probes associated with a label emitting a blue (B) signal").

-----continued in Supplemental Box-----

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

Form PCT/ISA/237 (Box No. VIII) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
Box IV-3 (Lack of Unity of Invention)

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore- labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 (Citations and Explanations)

Claims 88, 89 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of US 2010/0291557 A1 to LIVAK et al. (hereinafter "Livak").

As to claim 88, Cai teaches method, comprising: exposing a sample to a plurality of nucleic acid probes (para [0007]), and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes (para [0007]) formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences (para [0178]; "Super-resolution barcoding provides a useful single cell follow-up technique to high-throughput sequencing technologies"). Cai does not teach the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample. However, Livak teaches the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences (para [0062-0063]; "the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes specific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay") and for each of the nucleic acid probes determining binding of the target sequences of the nucleic acid and probes within the sample (para [0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Consequently, it would have been obvious for one of ordinary skill in the art to combine exposing a sample to a plurality of nucleic acid probes, at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences, all as taught by Cai, with the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, as taught by Livak, with determining binding of the target sequences of the nucleic acid and probes within the sample, also as taught by Livak, because it would have enabled a method of identifying and quantitating a mRNA population in a cell or group of cells.

As to claim 89, Cai further teaches creating codewords based on determination of the read sequences within the sample (para [0007], [0178]).

Claim 90 lacks an inventive step under PCT Article 33(3) as being obvious over Cai and Livak, in view of Hamady.

As to claim 90, Cai teaches codewords (para [0007]) and matching the codeword to a valid codeword (para [0007]). Neither Cai nor Livak teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches if no match is found, applying error correction to the codeword to form a valid codeword (para [0016], [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine codewords and matching the codeword to a valid codeword, as taught by Cai, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled because it would have enabled characterizing multiple isoforms of mRNA.

--------continued in next Supplemental Box-------

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 and previous Supplemental Box

Claims 166-171 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of Hamady, in further view of Livak, in final view of US 2004/0053300 A1 to SODERLUND et al (hereinafter "Soderlund").

As to claim 166, Cai teaches a method, comprising: associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords (para [0007]) wherein the codewords comprise a number of positions and values for each position (para [0007]). Cai does not teach the codewords form an error-checking and/or error-correcting code space. However, Hamady teaches that the codewords form an error-checking and/or error-correcting code space (para [0016]; [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Cai further teaches forming a plurality of nucleic acid probes (para [0007]) but does not teach that the plurality of probes comprise a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets. However, Livak teaches probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets (para [0062-0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Although Cai further teaches each of the probe pools corresponds to a specific value of a unique position within the codewords (para [0007]; [0162]- In experiments related to those previously described herein, the inventors targeted the YLR414c mRNA with3 sets of oligo probes labeled with different fluorophores"), Cai does not teach grouping the nucleic acid probes into a plurality of probe pools, exposing a sample to one of the probe pools; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions. However, Soderlund teaches grouping the nucleic acid probes into a plurality of probe pools (para [0049]; One or more soluble pools are provided with preset, but optional numbers, preferably varying between 10-500, more preferably between 50-400, most preferably between 100-300 soluble polynucleotide sequences, with distinct sizes allowing their accurate identification, e.g. by mass spectrometry. The soluble probes, which can be identified without any tracer tags, can alternatively be provided with tags"; para [0044]-" Preferred tracer tags are fluorochromes or fluorophors, especially such which have different wave-lengths of emission";. para [0033]-" The term "pool" means a subset or a library of soluble or solubilizable polynucleotide probes. Each pool comprises an optional defined number of polynucleotide probes. A convenient optional number is, for example, approximately 10 probes"; exposing a sample to one of the probe pools (para [0053]; In the basic embodiment of the present invention tracer or primer tagged DNA probes are allowed to hybridize with the preparation of RNA, which is to be analyzed. The analyte polynucleotide sequences or RNA analytes, present in the cell or tissue sample to be determined, are isolated by per se known methods"); Soderlund further teaches repeating the exposing step one or more times, using a different probe pool for at least some of the repetitions (para [0035]; However, because the nucleotide fragments are not characterized, some of the nucleotide fragments obtained by cleaving with various restriction enzymes may by chance have identical sizes. This means that the probes are only partly identifiable by their sizes. When repeating the tests, the redundancy is soon detected and such probes, which cannot be identified with 100% accuracy may be deleted or replaced with other more convenient probes"). Cai further teaches imaging a sample of probes (para [0141]; Imaging of the hybridized cells is carried out on automated fluorescence microscopes"). An artisan of ordinary skill in the art would have recognized that using pools or sets of probes would have enabled obtaining large amounts of information about a given sample, particularly if different probe pools were used in repetitive analysis, as taught by Soderlund. Consequently, it would have been obvious for one of ordinary skill in the art to combine associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, as taught by Cai, with the codewords form an error-checking and/or error-correcting code space, as taught by Hamady, with probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets, as taught by Livak, with grouping the nucleic acid probes into a plurality of probe pools, as taught by Soderlund, with exposing a sample to one of the probe pools; and repeating the exposing one or more times, using a different probe pool for at least some of the repetitions, again as taught by Soderlund, with using imaging technology to visualize the probes, as taught by Cai, because it would have enabled providing an expression profile of a sample.

As to claim 167, Cai further teaches combining the images of the sample into a combined image (para [0165]; "After many iterations of this cycle, a high resolution image can be reconstructed from all of the centroid fits in the individual images").

As to claim 168, Cai further teaches identifying foreground features corresponding to putative nucleic acid targets in the sample (para [0018]; enabling robust hybridization and identification by unique combinations of fluorophores").

As to claim 169, Cai further teaches determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample (para [0007], [0018], [0165]).

As to claim 170, Cai further teaches matching the determined codeword to a valid codeword wherein (para [0007]). Additionally, Hamady further teaches if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target (para [0016], [0098]).

As to claim 171, Cai further teaches imaging the sample using fluorescence microscopy (para [0141]; "Imaging of the hybridized cells is carried out on automated fluorescence microscopes").

Claims 1-3, 88-90, 166-171 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 February 2016 (04.02.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/018960 A1

(51) International Patent Classification:
C40B 30/04 (2006.01) *C12Q 1/68* (2006.01)
C07H 21/02 (2006.01)

(21) International Application Number:
PCT/US2015/042556

(22) International Filing Date:
29 July 2015 (29.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/031,062 30 July 2014 (30.07.2014) US
62/050,636 15 September 2014 (15.09.2014) US
62/142,653 3 April 2015 (03.04.2015) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **ZHUANG, Xiaowei**; 5 Luongo Farm Lane, Lexington, MA 02421 (US). **CHEN, Kok-Hao**; 37 Jalan Lepan, Singapore 537961 (SG). **BOETTIGER, Alistair**; 1550 Cambridge Street, Apt. #2, Cambridge, MA 02139 (US). **MOFFITT, Jeffrey, R.**; 175 Beacon Street, Apt. #108, Somerville, MA 02143 (US). **WANG, Siyuan**; 28 Fernald Drive, Apt. #31, Cambridge, MA 02138 (US).

(74) Agent: **CHEN, Tani**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

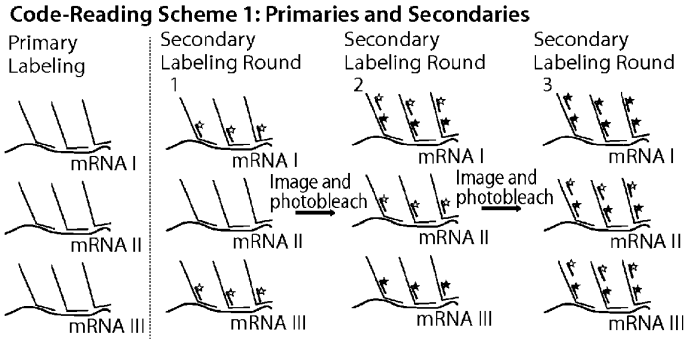


Fig. 1B

(57) Abstract: The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

WO 2016/018960 A1

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first

secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Fig. 12 shows schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16B show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention; and

Figs. 19A-9D illustrate decoding and error assessment, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:
GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA
TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG
GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;

SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;

SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;

SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and

SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the

cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences.

For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize

a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was

shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that

allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%,

at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell

may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40,

at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For

instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at

least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments,

the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting

code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n …) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this

number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or

more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about 500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about

0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification

and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic

acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or “quantum dots”, fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities

are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another

set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light

having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be

connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate,

Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso

Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate

control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled "Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate,

polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those

sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three

Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

In addition, incorporated herein by reference in their entireties are U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and a PCT application filed on even date herewith, entitled "Probe Library Construction," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the

bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs

computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps (meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 ^{o}C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10
B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected

while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 ^{o}C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 ^{o}C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final

1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics

setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0]. Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors. The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not

achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as N increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper

symbols). With realistic error rates per hybridization (measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the *N* unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these *N* readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via *N* rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse

transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in *N* rounds of imaging. Each RNA species is encoded with a *N*-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (*N*) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of *N* different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the *N* readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. *N* subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb *N*). The bound probes were inactivated by photobleaching between successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA

species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, …), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction. To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro* transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes

containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Fig. 20). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control

words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Fig. 20) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the average copy number and the copy number distribution across many cells agreed quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range

additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-value of 1×10^{-53}. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads (FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of 3×10^{-39}.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4

code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e., a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification

control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from ~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Fig. 20 shows two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated

or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed, this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1, PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by

increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient $r = 0.76$; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data ($r = 0.65$; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient is 0.76 with a p-value of 3×10^{-133}. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of 3×10^{-33}. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of 9×10^{-30}. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene

experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes). The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the

140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACCG (SEQ ID NO: 20)	TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC (SEQ ID NO: 23)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of

homology) and for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1 with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 ^{o}C; there is no cross hybridization targets with melting temperatures greater than 72 ^{o}C; there is no predicted internal secondary structures with melting temperatures greater than 76 ^{o}C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using

BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 ^{o}C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 ^{o}C for 45 minutes, and the reverse transcriptase was inactivated at 85 ^{o}C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG (SEQ ID NO: 40). A common priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG (SEQ ID NO: 20).

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 ^{o}C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 µL of encoding hybridization buffer (see below). Probes were stored at -20 ^{o}C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to

confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 ^{o}C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 ^{o}C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 ^{o}C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5 minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged

immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope

backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold

required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1 pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all

subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above) was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 ^{o}C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 ^{o}C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible code words, the calling rate, and the misidentification rate on *N*. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong

species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00…0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced by this approximation is negligible. For any given word with *m* '1's and *N*–*m* '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \quad (1)$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \quad (2)$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with *m* '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the

fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m} + m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m} + (N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the m '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the N-m '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of m. The weight for each term was determined from the number of words that contain m '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with m '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^{\,i}p_0^{\,4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)} + \\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^{\,i}p_0^{\,3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)} + \\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^{\,i}p_0^{\,5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)} +
\end{aligned} \quad (5)$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have m '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did

not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}&\binom{4}{2}\binom{N-4}{2}p_1^2p_0^2(1-p_1)^2(1-p_0)^{N-6}+\\&\binom{4}{1}\binom{N-4}{2}p_1^1p_0^2(1-p_1)^3(1-p_0)^{N-6}+\binom{4}{2}\binom{N-4}{1}p_1^2p_0^1(1-p_1)^2(1-p_0)^{N-5}\\&\binom{4}{2}\binom{N-4}{3}p_1^2p_0^3(1-p_1)^2(1-p_0)^{N-7}+\binom{4}{3}\binom{N-4}{2}p_1^3p_0^2(1-p_1)^1(1-p_0)^{N-6}\end{aligned} \quad (6)$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be A, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \quad (7)$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in

order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and

optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

1. A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2. The method of claim 1, comprising exposing the sample to at least 5 different nucleic acid probes.

3. The method of any one of claims 1 or 2, comprising exposing the sample to at least 10 different nucleic acid probes.

4. The method of any one of claims 1-3, comprising exposing the sample to at least 100 different nucleic acid probes.

5. The method of any one of claims 1-4, comprising exposing the sample to a plurality of nucleic acid probes simultaneously.

6. The method of any one of claims 1-5, comprising exposing the sample to a plurality of nucleic acid probes sequentially.

7. The method of claims 1-6, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8. The method of claim 7, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of RNA species in the sample.

9. The method of claim 7 or 8, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of DNA sequences in the sample.

10. The method of any one of claims 1-9, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11. The method of claim 10, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

12. The method of any one of claims 10 or 11, wherein the target sequence is substantially complementary to a nucleic acid sequence encoding a protein.

13. The method of any one of claims 10-12, wherein the target sequence is substantially complementary to messenger RNA (mRNA).

14. The method of any one of claims 10-13, wherein the plurality of nucleic acid probes comprises at least 8 possible read sequences.

15. The method of any one of claims 10-14, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

16. The method of any one of claims 10-15, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

17. The method of any one of claims 10-16, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

18. The method of any one of claims 10-17, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19. The method of any one of claims 10-18, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

20. The method of any one of claims 10-19, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

21. The method of any one of claims 10-20, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

22. The method of any one of claims 10-21, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

23. The method of any one of claims 10-22, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

24. The method of any one of claims 10-23, wherein at least some of the plurality of nucleic acid probes comprise no more than 3 read sequences.

25. The method of any one of claims 10-24, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

26. The method of claims 1-25, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27. The method of claim 26, wherein the first signaling entity is fluorescent.

28. The method of any one of claims 26 or 27, wherein the first signaling entity is a protein.

29. The method of any one of claims 26-28, wherein the first signaling entity is a dye.

30. The method of any one of claims 26-29, wherein the first signaling entity is a nanoparticle.

31. The method of any one of claims 26-30, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32. The method of claim 31, wherein the first signaling entity and the second signaling entity are identical.

33. The method of claim 31, wherein the first signaling entity and the second signaling entity are not identical.

34. The method of any one of claims 31-33, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35. The method of claim 34, comprising inactivating the first signaling entity by photobleaching at least some of the first signaling entity.

36. The method of any one of claims 34 or 35, comprising inactivating the first signaling entity by chemically bleaching at least some of the first signaling entity.

37. The method of any one of claims 34-36, comprising inactivating the first signaling entity by exposing the first signaling entity to a reactant able to alter the structure of the signaling entity.

38. The method of any one of claims 34-37, comprising inactivating the first signaling entity by removing at least some of the first signaling entity.

39. The method of any one of claims 34-38, comprising inactivating the first signaling entity by dissociating the first signaling entity from the first secondary probe.

40. The method of any one of claims 34-39, comprising inactivating the first signaling entity by dissociating the first secondary probe that contains the first signaling entity from the sample.

41. The method of any one of the claims 34-40, comprising inactivating the first signaling entity by chemically cleaving it from the first secondary probe.

42. The method of any one of the claims 34-41, comprising inactivating the first signaling entity by enzymatically cleaving it from the first secondary probe.

43. The method of any one of claims 34-42, comprising inactivating the first signaling entity by exposing the signaling entity or the first secondary probe to an enzyme.

44. The method of any one of claims 1-43, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

45. The method of any one of claims 1-44, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

46. The method of any one of claims 44 or 45, comprising determining the centroid using a maximum likelihood algorithm.

47. The method of any one of claims 44-46, comprising determining the centroid using a least squares algorithm.

48. The method of any one of claims 44-47, comprising determining the centroid using a Bayesian algorithm.

49. The method of any one of claims 44-48, comprising determining the centroid using a compressed sensing algorithm.

50. The method of any one of claims 1-49, further comprising determining a confidence level for the identified nucleic acid target.

51. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having one or more one-bit errors to the codeword.

52. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having exactly one one-bit error to the codeword.

53. The method of claims 1-52, wherein at least some of the plurality of nucleic acid probes comprise DNA.

54. The method of any one of claims 1-53, wherein at least some of the plurality of nucleic acid probes comprise RNA.

55. The method of any one of claims 1-54, wherein at least some of the plurality of nucleic acid probes comprise PNA.

56. The method of any one of claims 1-55, wherein at least some of the plurality of nucleic acid probes comprise LNA.

57. The method of any one of claims 1-56, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

58. The method of any one of claims 1-57, wherein at least some of the plurality of nucleic acids probes are configured to bind to a nucleic acid within the sample.

59. The method of any one of claims 1-58, wherein at least some of the binding of the nucleic acid probes to a target within the sample is specific binding.

60. The method of any one of claims 1-59, wherein at least some of the binding of the nucleic acid probes to a target within the sample is via Watson-Crick base pairing.

61. The method of any one of claims 1-60, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

62. The method of any one of claims 1-61, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

63. The method of any one of claims 1-62, wherein at least some of the plurality of nucleic acid probes are configured to bind to mRNA.

64. The method of any one of claims 1-63, wherein at least some of the plurality of nucleic acid probes are configured to bind to transfer RNA (tRNA).

65. The method of any one of claims 1-64, wherein at least some of the plurality of nucleic acid probes are configured to bind to ribosomal RNA (rRNA).

66. The method of any one of claims 1-65, wherein at least some of the plurality of nucleic acid probes are configured to bind to lncRNA.

67. The method of any one of claims 1-66, wherein at least some of the plurality of nucleic acid probes are configured to bind to snoRNA.

68. The method of any one of claims 1-67, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

69. The method of any one of claims 1-68, wherein the plurality of nucleic acids probes is configured to bind to DNA.

70. The method of any one of claims 1-69, wherein at least some of the plurality of nucleic acid probes are configured to bind to genomic DNA.

71. The method of any one of claims 1-70, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

72. The method of any one of claims 1-71, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm.

73. The method of any one of claims 1-72, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 80 nm.

74. The method of any one of claims 1-73, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 50 nm.

75. The method of any one of claims 1-74, wherein the sample comprises a cell.

76. The method of claim 75, wherein the cell is a human cell.

77. The method of any one of claims 76 or 77, wherein the cell is fixed.

78. The method of any one of claims 1-77, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

79. The method of any one of claims 1-78, comprising determining binding of the nucleic acid probes using an optical imaging technique.

80. The method of any one of claims 1-79, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

81. The method of any one of claims 1-80, comprising determining binding of the nucleic acid probes using a multi-color fluorescence imaging technique.

82. The method of any one of claims 1-81, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

83. The method of claim 82, comprising determining binding of the nucleic acid probes using stochastic optical reconstruction microscopy (STORM).

84. The method of any one of claims 82 or 83, comprising determining binding of the nucleic acid probes using photoactivated localization microscopy (PALM) or fluorescence photoactivation localization microscopy (FPALM).

85. The method of any one of claims 82-84, comprising determining binding of the nucleic acid probes using stimulated emission depletion microscopy (STED).

86. The method of any one of claims 82-85, comprising determining binding of the nucleic acid probes using structured illumination microscopy (SIM).

87. The method of any one of claims 82-86, comprising determining binding of the nucleic acid probes using reversible saturated optically linear fluorescence transition (RESOLFT) microscopy.

88. A method, comprising:

exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and

for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

89. The method of claim 88, further comprising:

creating codewords based on determination of the read sequences within the sample.

90. The method of claim 89, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

91. A method, comprising:

exposing a sample to a plurality of primary nucleic acid probes;

exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample;

creating codewords based on fluorescence of the secondary nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

92. A method, comprising:

exposing a plurality of primary nucleic acid probes to a sample; and

exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample, wherein at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

93. The method of claim 92, further comprising:

creating codewords based on fluorescence of the secondary nucleic acid probes.

94. The method of claim 93, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

95. A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

96. The composition of claim 95, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

97. The composition of claim 96, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 3.

98. The composition of any one of claims 96 or 97, wherein the code is a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, or a Hamming(127, 120) code.

99. The composition of any one of claims 95-98, wherein the code is a SECDED code.

100. The composition of claim 99, wherein the code is a SECDED(8,4) code.

101. The composition of claim 99, wherein the code is a SECDED(16,4) code.

102. The composition of claim 99, wherein the code is a SECDED(16, 11) code, a SECDED (22, 16) code, a SECDED (39, 32) code, or a SECDED (72, 64) code.

103. The composition of any one of claims 99-102, wherein only codewords with a constant number of ones are used.

104. The composition of claim 99, wherein the code is MHD4 code.

105. The composition of claim 99, wherein the code is MHD2 code.

106. The composition of any one of claims 95-105, wherein the plurality of nucleic acid probes comprises no more than 100 possible read sequences.

107. The composition of any one of claims 95-106, wherein the plurality of nucleic acid probes comprises no more than 64 possible read sequences.

108. The composition of any one of claims 95-107, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

109. The composition of any one of claims 95-108, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

110. The composition of any one of claims 95-109, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

111. The composition of any one of claims 95-110, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

112. The composition of any one of claims 95-111, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

113. The composition of any one of claims 95-112, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

114. The composition of any one of claims 95-113, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

115. The composition of any one of claims 95-114, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

116. The composition of any one of claims 95-115, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

117. The composition of any one of claims 95-116, wherein at least some of the plurality of nucleic acid probes comprise DNA.

118. The composition of any one of claims 95-117, wherein at least some of the plurality of nucleic acid probes comprise RNA.

119. The composition of any one of claims 95-118, wherein at least some of the plurality of nucleic acid probes comprise PNA.

120. The composition of any one of claims 95-119, wherein at least some of the plurality of nucleic acid probes comprise LNA.

121. The composition of any one of claims 95-120, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

122. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

123. The method of claim 122, wherein the code space has a Hamming distance of at least 2.

124. The method of any one of claims 122 or 123, wherein the plurality of targets comprise RNA.

125. The method of any one of claims 122-124, wherein the plurality of targets comprise DNA.

126. A method, comprising:

exposing nucleic acid probes formed using the method of any one of claims 122-125 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets

within the sample.

127. The method of claim 126, wherein at least some of the plurality of secondary nucleic probes is able to bind to a read sequence of a nucleic acid probe.

128. The method of claim 127, wherein at least some of the plurality of secondary nucleic probes is able to specifically bind to only one read sequence.

129. The method of claim 125-128, wherein if no match is found between a valid codeword and the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

130. The method of claim 125-129, wherein at least some of the secondary nucleic acid probes each comprise a signaling entity.

131. The method of claim 130, comprising imaging the signaling entity to determine binding of the plurality of secondary nucleic probes within the sample.

132. The method of claim 131, comprising imaging the signaling entity using a super-resolution imaging technique.

133. The method of any one of claims 131 or 132, further comprising applying drift correction to the images.

134. The method of any one of claims 131-133, further comprising applying physical registration to the images.

135. The method of claim 125-134, comprising determining binding of the plurality of secondary nucleic probes within the sample at a resolution better than 100 nm.

136. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-

correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

137. The method of claim 136, where in the codewords comprise a number of positions that is less than the number of targets.

138. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

139. The method of claim 138, wherein the codewords form an error-checking space.

140. The method of any one of claims 138 or 139, wherein the codewords form an error-correcting code space.

141. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each

group of nucleic acid probes correspond to at least one common value of a position within the codewords.

142. The method of any one of claims 138-141, wherein the groups are applied to the sample sequentially using fluidic devices.

143. The method of any one of claims 138-142, wherein each group of nucleic acid probes contains a group identification sequence distinguishable from the group identification sequence of the other groups.

144. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code;

associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample;

forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences;

forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences;

exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample;

exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

145. The method of claim 144, wherein the wherein the read sequences are taken from a set of orthogonal sequences, which have a homology of less than 15 basepairs with

one another and with the nucleic acid species in a sample.

146. The method of any one of claims 144 or 145, comprising combining the images of the sample into a combined image.

147. The method of claim 146, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

148. The method of any one of claims 144-147, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

149. The method of any one of claims 144-148, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

150. The method of any one of claims 144-149, comprising determining the centroid using a maximum likelihood algorithm.

151. The method of any one of claims 144-150, comprising determining the centroid using a least squares algorithm.

152. The method of any one of claims 144-151, comprising determining the centroid using a Bayesian algorithm.

153 The method of any one of claims 144-152, comprising determining the centroid using a compressed sensing algorithm.

154. The method of any one of claims 144-153, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

155. The method of claim 154, comprising matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a

valid codeword that correspond to a nucleic acid target

156. The method of any one of claims 144-155, comprising imaging the sample using fluorescence microscopy.

157. The method of any one of claims 144-156, comprising imaging the sample at a resolution better than 500 nm.

158. The method of any one of claims 144-157, comprising imaging the sample at a resolution better than 300 nm.

159. The method of any one of claims 144-158, comprising imaging the sample at a resolution better than 100 nm.

160. The method of any one of claims 144-159, comprising imaging the sample at a resolution better than 80 nm.

161. The method of any one of claims 144-160, comprising imaging the sample at a resolution better than 50 nm.

162. The method of any one of claims 144-161, comprising imaging the sample at a resolution better than 20 nm.

163. The method of any one of claims 144-162, comprising imaging the sample using a super-resolution fluorescence imaging technique.

164. The method of any one of claims 144-163, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

165. The method of any one of claims 144-164, where in the codewords comprise a number of positions that is less than the number of targets.

166. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets;

grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords;

exposing a sample to one of the probe pools;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

167. The method of claim 166, comprising combining the images of the sample into a combined image.

168. The method of claim 167, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

169. The method of claim 168, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

170. The method of claim 169, comprising matching the determined codeword to a valid codeword wherein, if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target.

171. The method of any one of claims 166-170, comprising imaging the sample using fluorescence microscopy.

172. The method of any one of claims 166-171, comprising imaging the sample at a resolution better than 300 nm.

173. The method of any one of claims 166-172, comprising imaging the sample at a resolution better than 100 nm.

174. The method of any one of claims 166-173, comprising imaging the sample using a super-resolution fluorescence imaging technique.

175. The method of any one of claims 166-174, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

176. The method of any one of claims 166-175, wherein the codewords comprise a number of positions that is less than the number of targets.

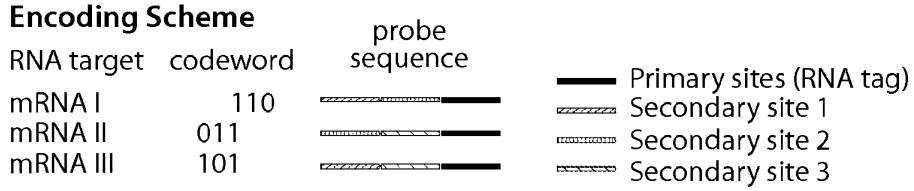


Fig. 1A

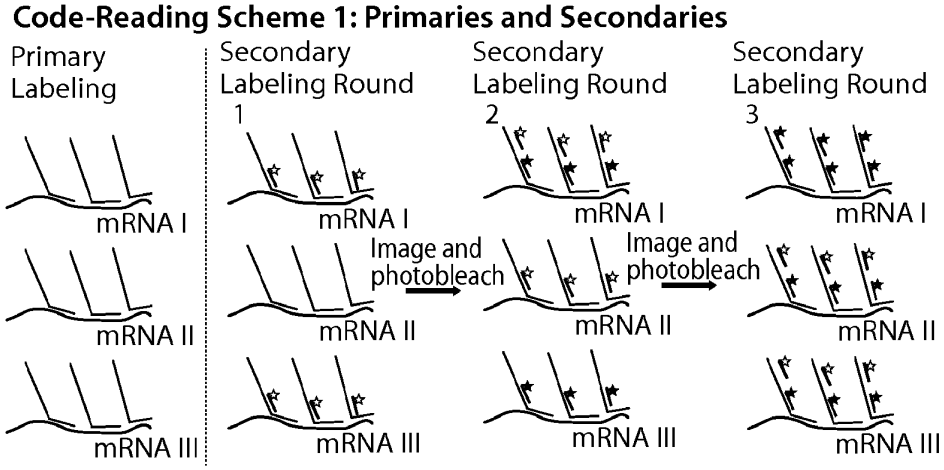


Fig. 1B

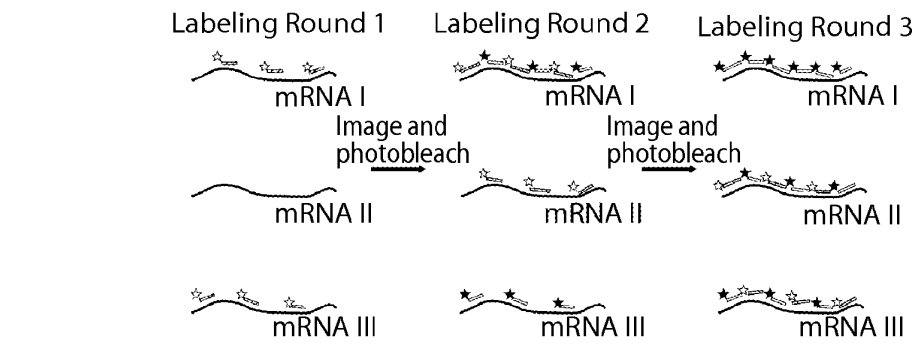


Fig. 1C

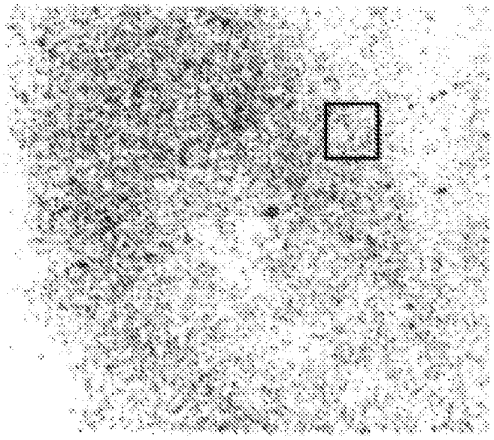


Fig. 2A

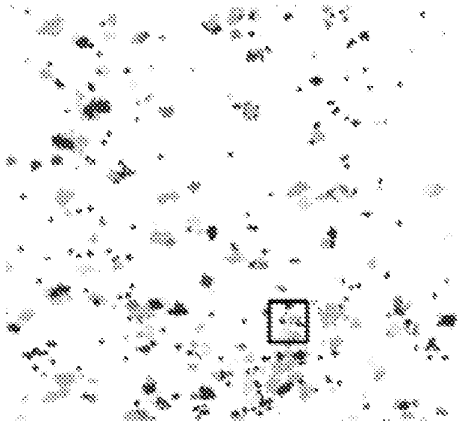


Fig. 2B



[0 1 0 1 1 1 0 0]

Fig. 2C

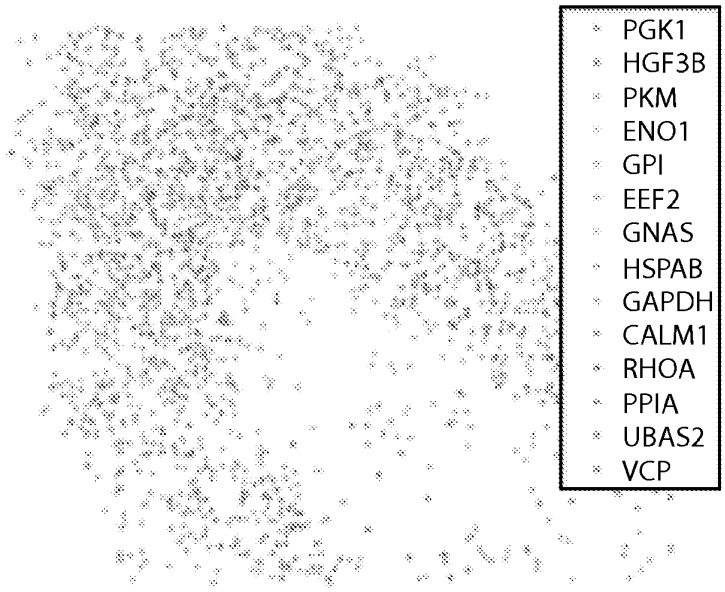

PGK1
HGF3B
PKM
ENO1
GPI
EEF2
GNAS
HSPAB
GAPDH
CALM1
RHOA
PPIA
UBAS2
VCP

Fig. 2D

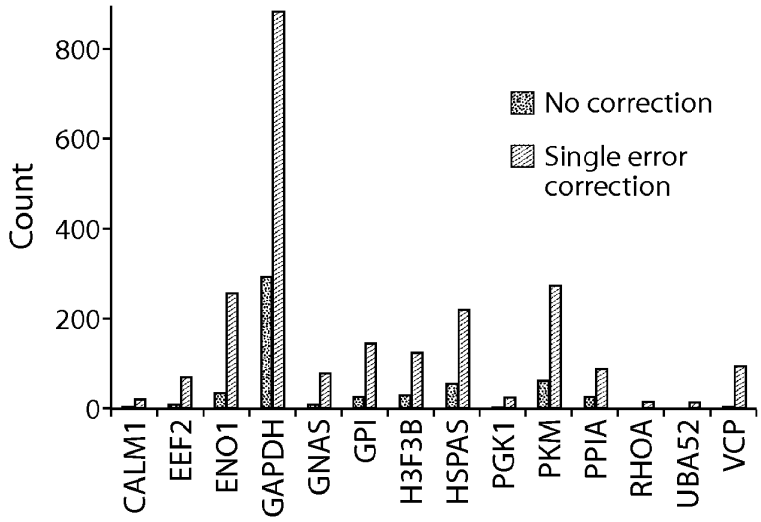

Count
800
600
400
200
0
No correction
Single error correction
CALM1
EEF2
ENO1
GAPDH
GNAS
GPI
H3F3B
HSPAS
PGK1
PKM
PPIA
RHOA
UBA52
VCP

Fig. 2E

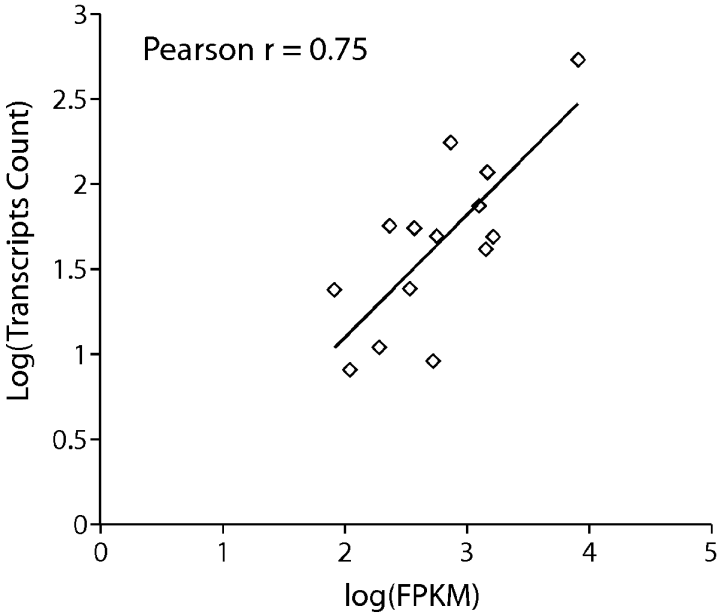

Pearson r = 0.75
Log(Transcripts Count)
log(FPKM)
0
0.5
1
1.5
2
2.5
3
0
1
2
3
4
5

Fig. 2F

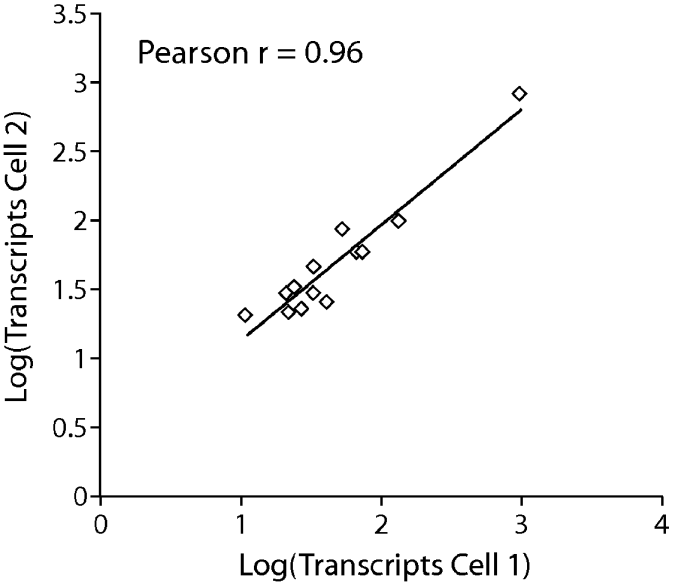

Pearson r = 0.96
Log(Transcripts Cell 2)
Log(Transcripts Cell 1)
0
0.5
1
1.5
2
2.5
3
3.5
0
1
2
3
4

Fig. 2G

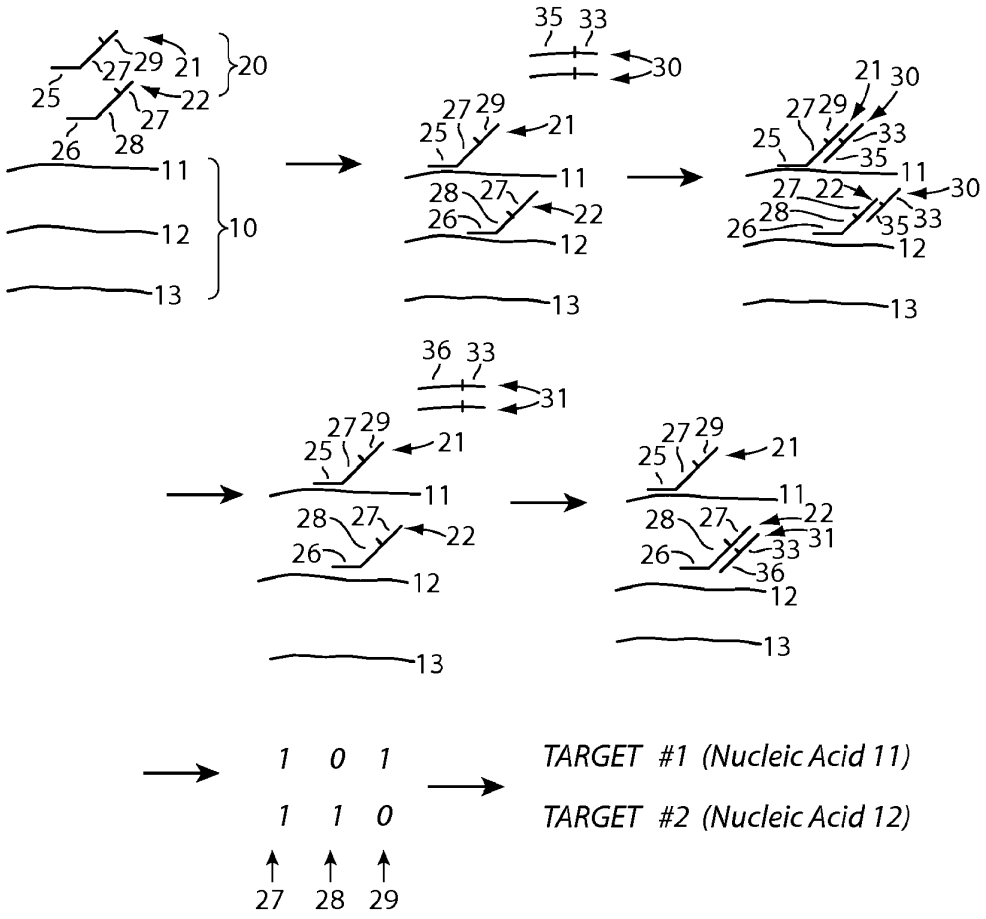

20
10
11
12
13
21
22
25
26
27
28
29
30
31
33
35
36
1 0 1
1 1 0
TARGET #1 (Nucleic Acid 11)
TARGET #2 (Nucleic Acid 12)

Fig. 3A

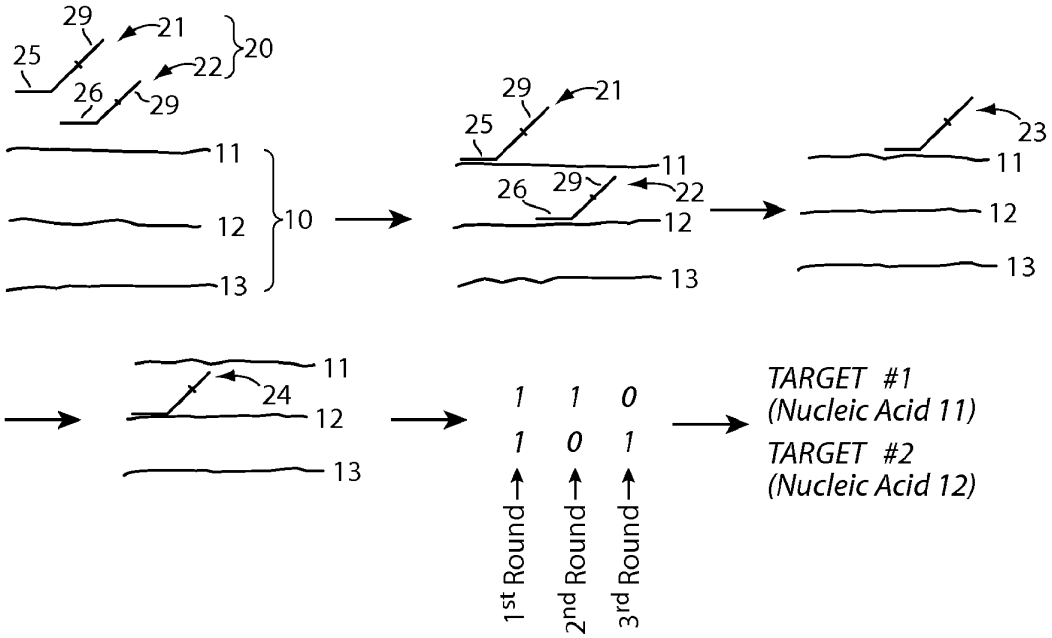


Fig. 3B

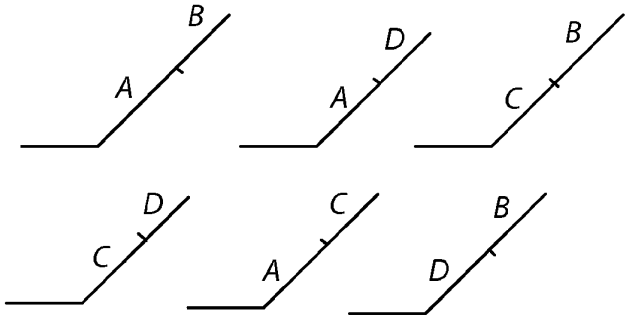


Fig. 4A

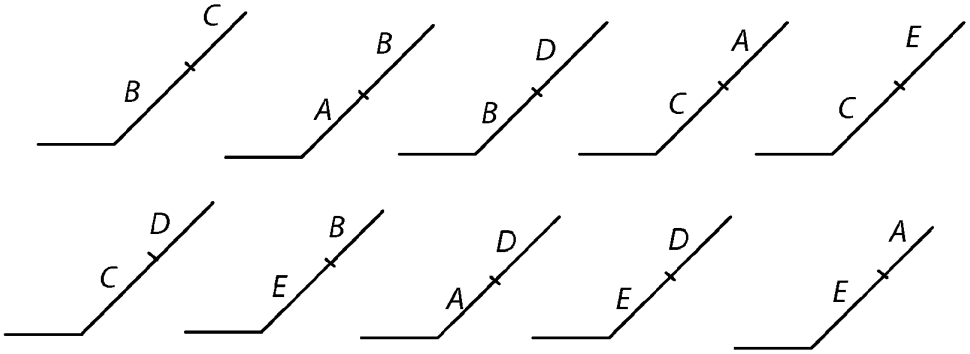


Fig. 4B

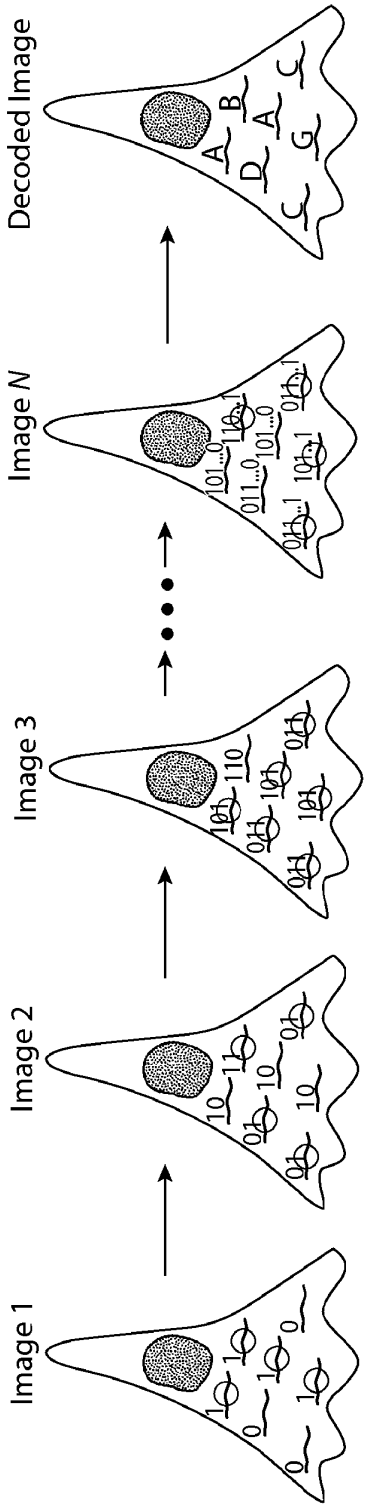


Fig. 5A

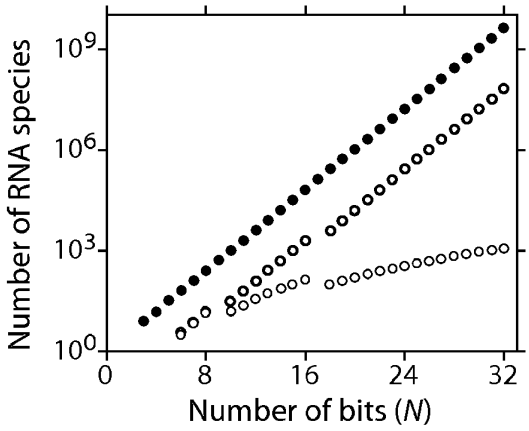


Fig. 5B

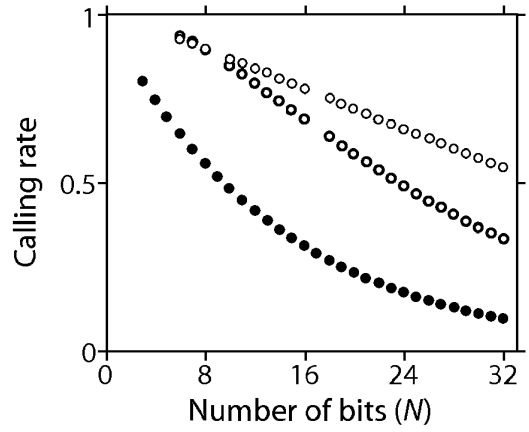


Fig. 5C

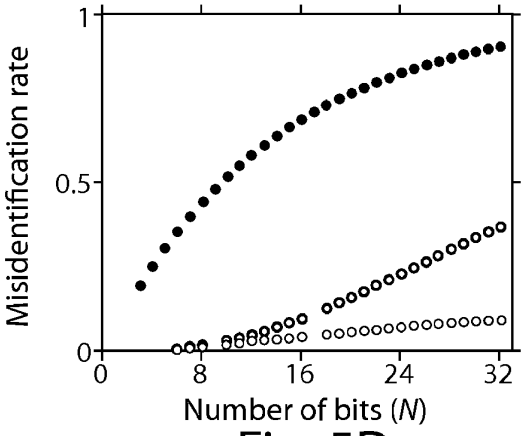


Fig. 5D

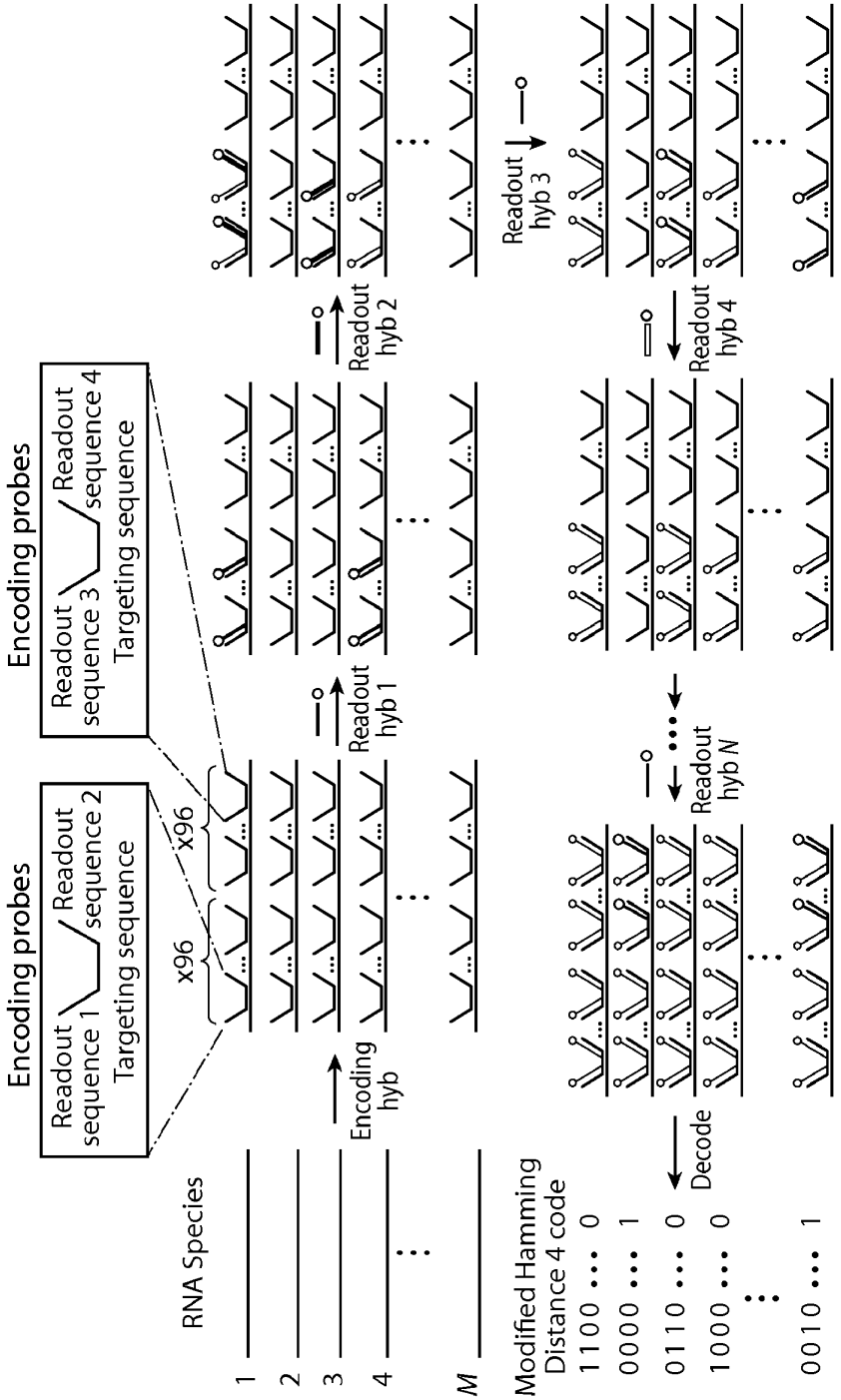

Encoding probes
Readout sequence 1
Readout sequence 2
Targeting sequence
Encoding probes
Readout sequence 3
Readout sequence 4
Targeting sequence
x96
x96
RNA Species
1
2
3
4
M
Encoding hyb
Readout hyb 1
Readout hyb 2
Readout hyb 3
Readout hyb 4
Readout hyb N
Modified Hamming Distance 4 code
1100 ... 0
0000 ... 1
0110 ... 0
1000 ... 0
0010 ... 1
Decode

Fig. 5E

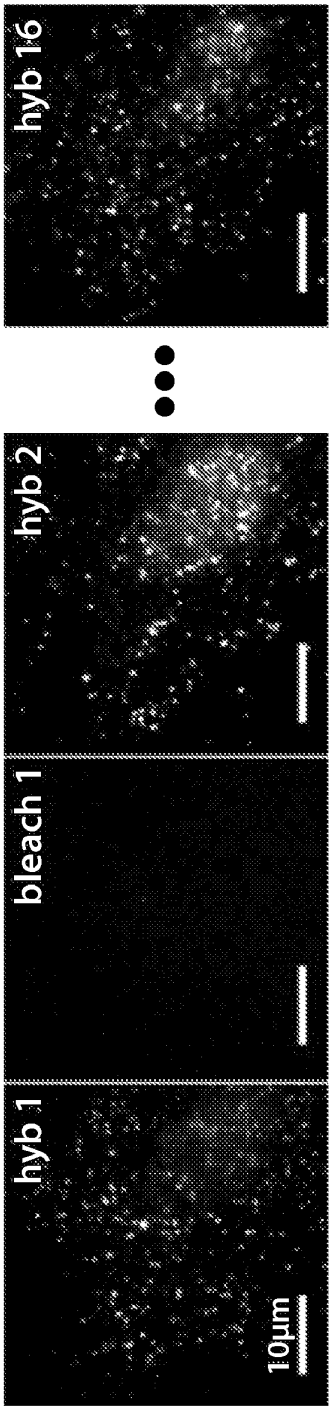


Fig. 6A

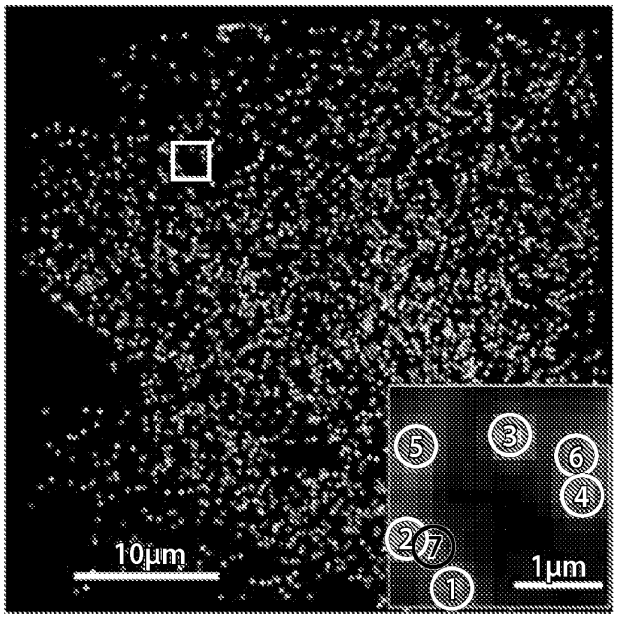

10µm
1µm
1
2
3
4
5
6
7

Fig. 6B

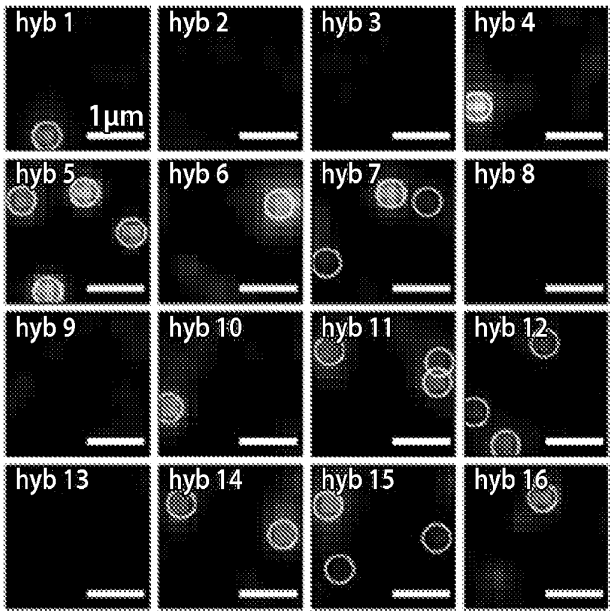

hyb 1
1µm
hyb 2
hyb 3
hyb 4
hyb 5
hyb 6
hyb 7
hyb 8
hyb 9
hyb 10
hyb 11
hyb 12
hyb 13
hyb 14
hyb 15
hyb 16

Fig. 6C

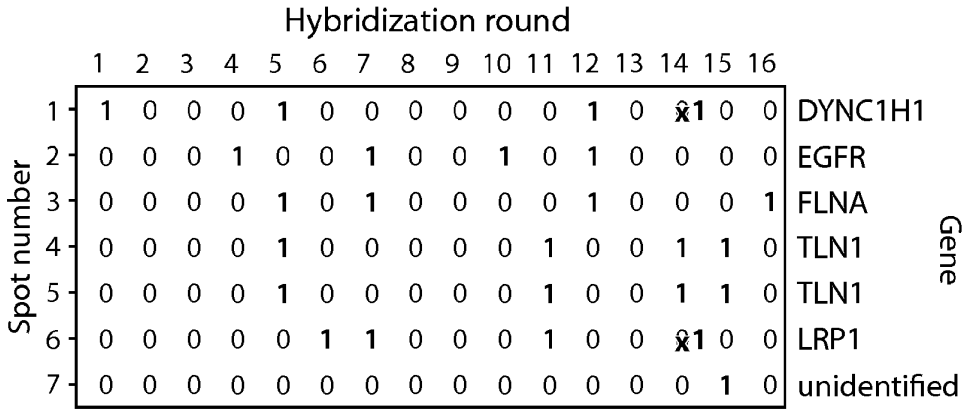

Hybridization round
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Spot number
1 1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0 DYNC1H1
2 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 EGFR
3 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 FLNA
4 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 TLN1
5 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 TLN1
6 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 LRP1
7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 unidentified
Gene

Fig. 6D

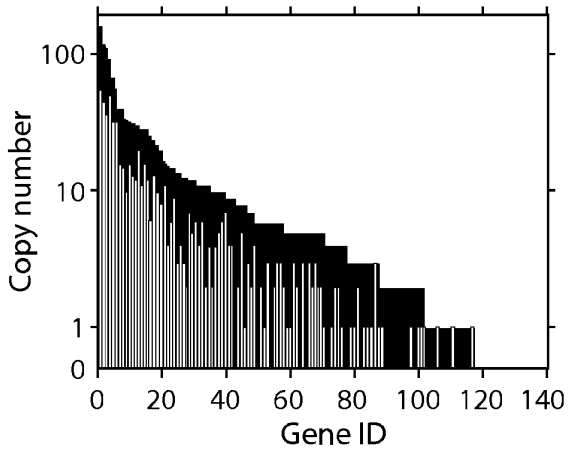

Copy number
100
10
1
0
0 20 40 60 80 100 120 140
Gene ID

Fig. 6E

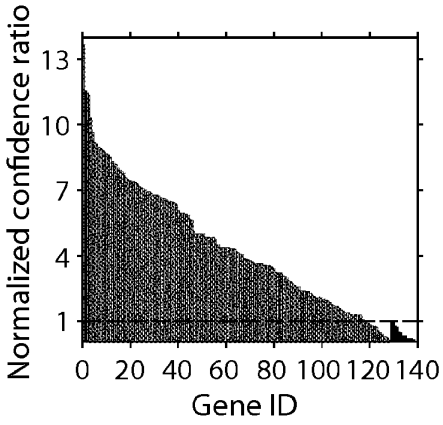


Fig. 6F

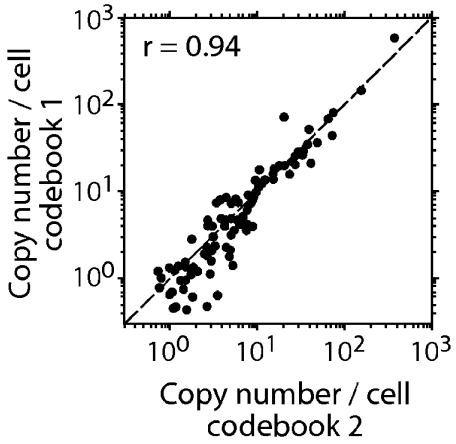


Fig. 6G

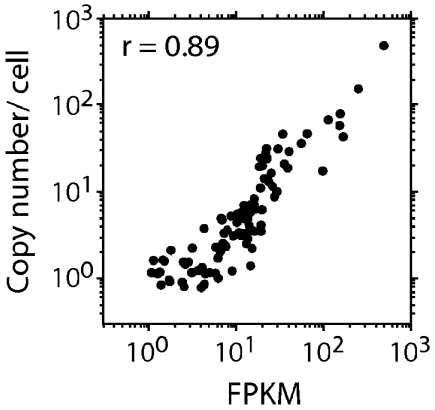


Fig. 6H

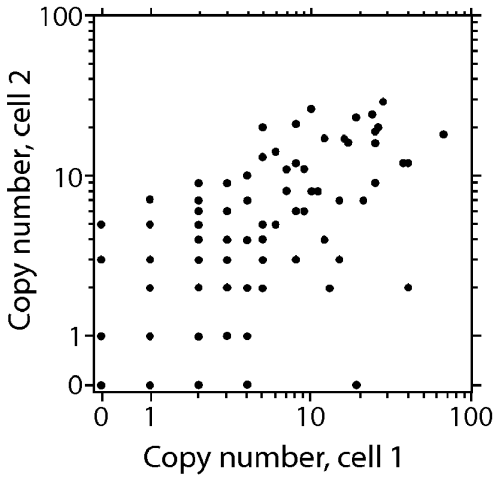

100
10
1
0
0
1
10
100
Copy number, cell 2
Copy number, cell 1

Fig. 7A

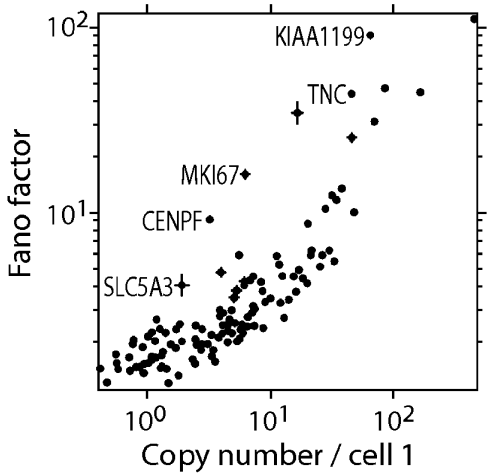

10^2
10^1
10^0
10^1
10^2
KIAA1199
TNC
MKI67
CENPF
SLC5A3
Fano factor
Copy number / cell 1

Fig. 7B

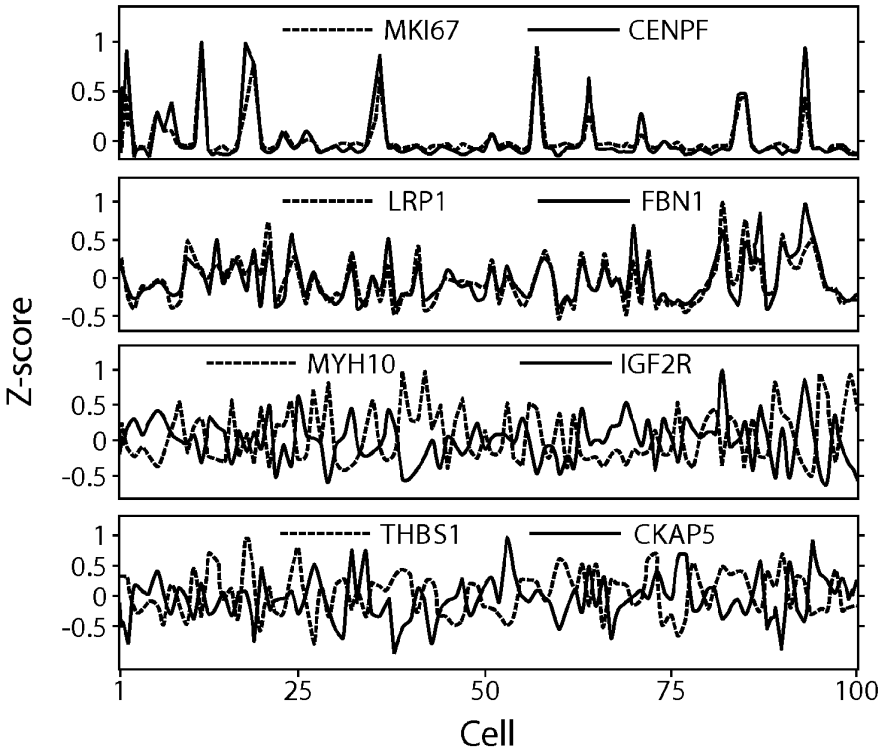

MKI67
CENPF
LRP1
FBN1
MYH10
IGF2R
THBS1
CKAP5
Z-score
1
0.5
0
-0.5
1
25
50
75
100
Cell

Fig. 7C

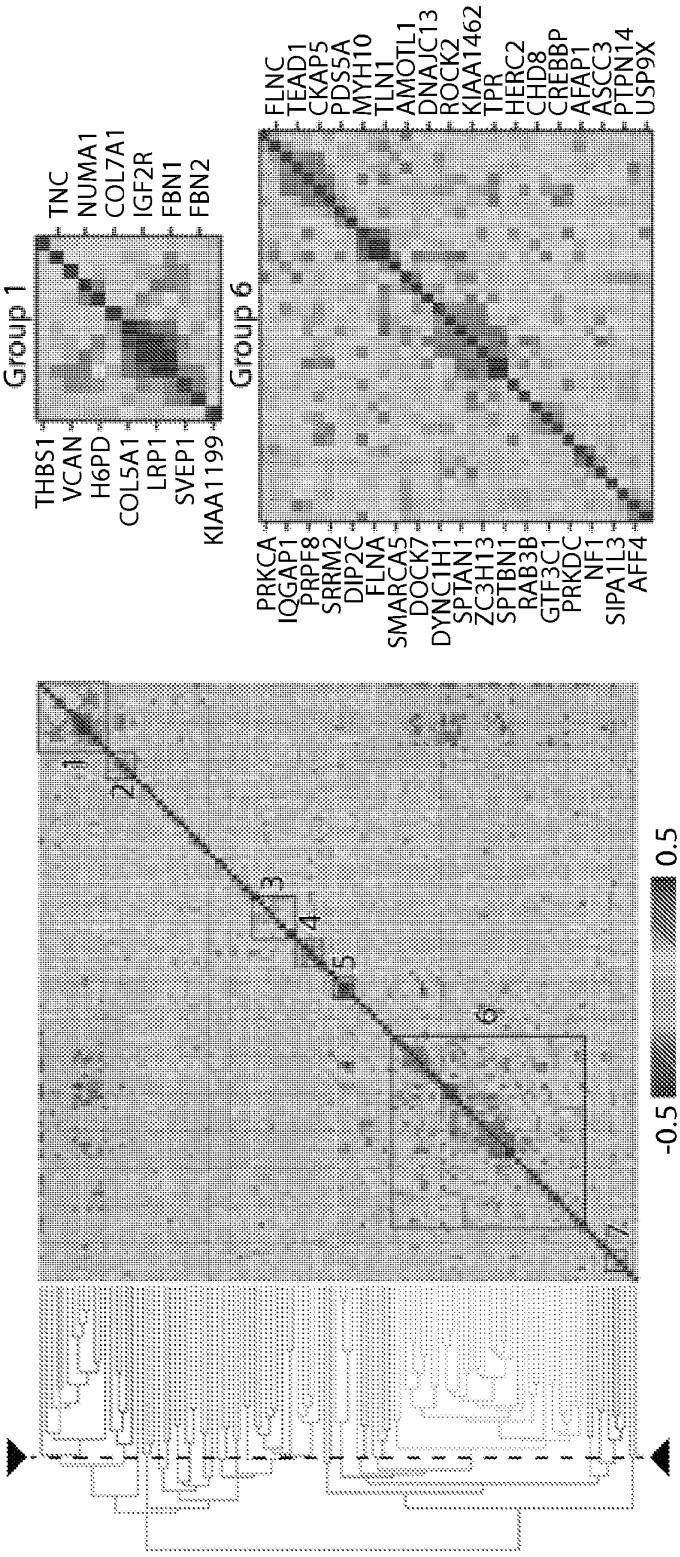

Group 1
TNC
NUMA1
COL7A1
IGF2R
FBN1
FBN2
THBS1
VCAN
H6PD
COL5A1
LRP1
SVEP1
KIAA1199
Group 6
FLNC
TEAD1
CKAP5
PDS5A
MYH10
TLN1
AMOTL1
DNAJC13
ROCK2
KIAA1462
TPR
HERC2
CHD8
CREBBP
AFAP1
ASCC3
PTPN14
USP9X
PRKCA
IQGAP1
PRPF8
SRRM2
DIP2C
FLNA
SMARCA5
DOCK7
DYNC1H1
SPTAN1
ZC3H13
SPTBN1
RAB3B
GTF3C1
PRKDC
NF1
SIPA1L3
AFF4
1
2
3
4
5
6
7
-0.5
0.5

Fig. 7D

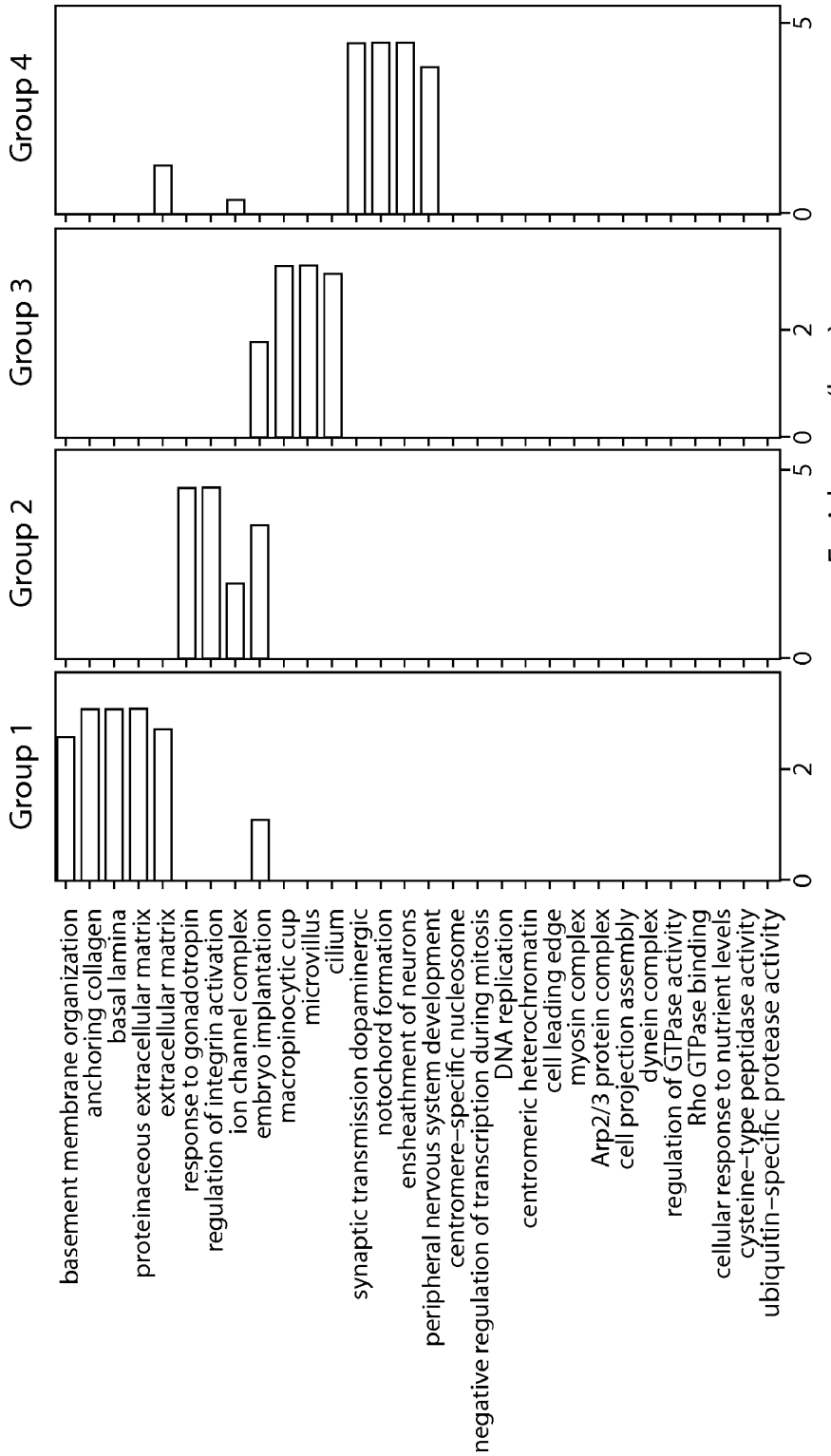

Group 1
Group 2
Group 3
Group 4
0
2
0
5
0
2
0
5
Enrichment (log2)
basement membrane organization
anchoring collagen
basal lamina
proteinaceous extracellular matrix
extracellular matrix
response to gonadotropin
regulation of integrin activation
ion channel complex
embryo implantation
macropinocytic cup
microvillus
cilium
synaptic transmission dopaminergic
notochord formation
ensheathment of neurons
peripheral nervous system development
centromere−specific nucleosome
negative regulation of transcription during mitosis
DNA replication
centromeric heterochromatin
cell leading edge
myosin complex
Arp2/3 protein complex
cell projection assembly
dynein complex
regulation of GTPase activity
Rho GTPase binding
cellular response to nutrient levels
cysteine−type peptidase activity
ubiquitin−specific protease activity

Fig. 7E

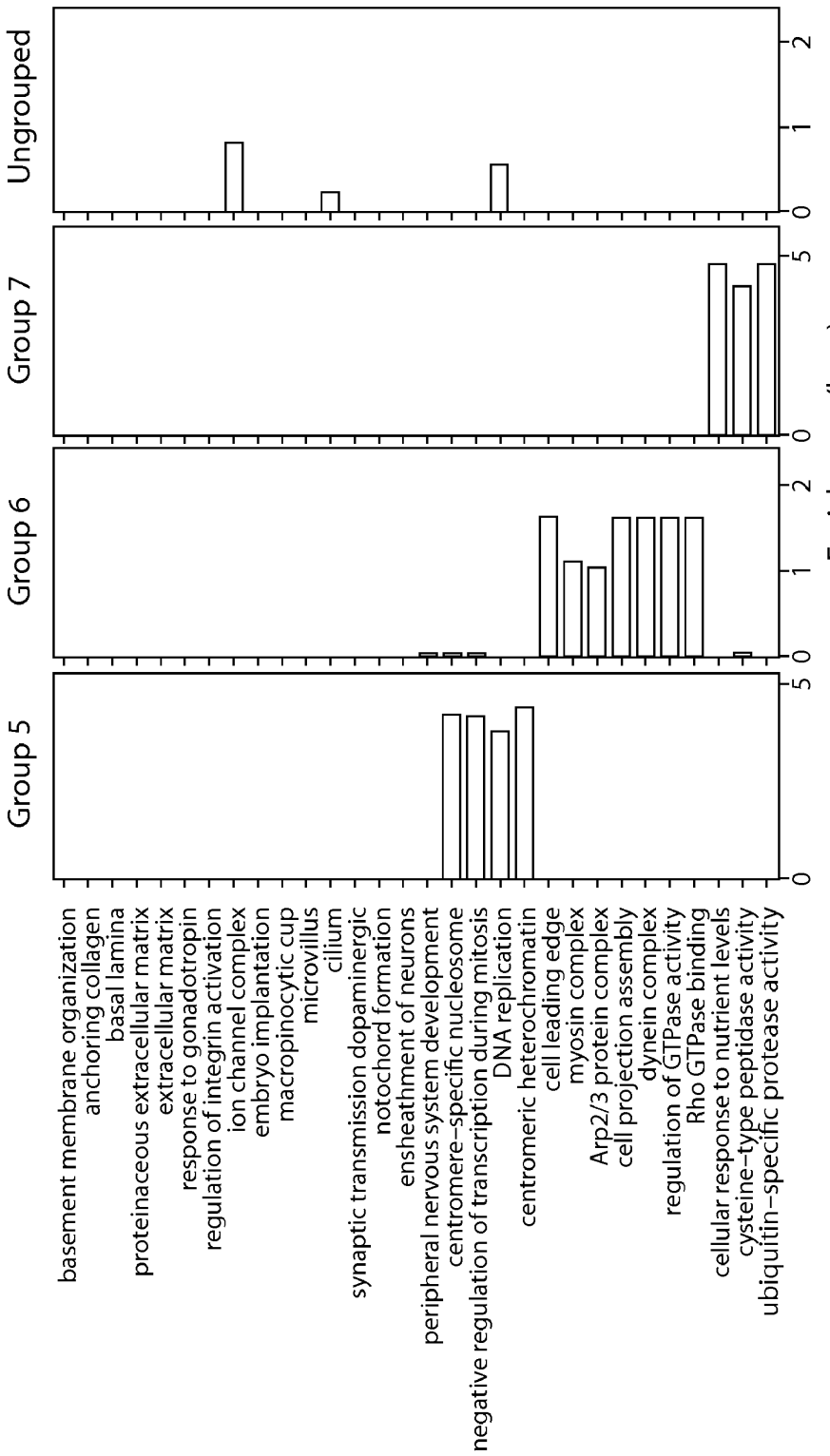


Fig. 7E Continued

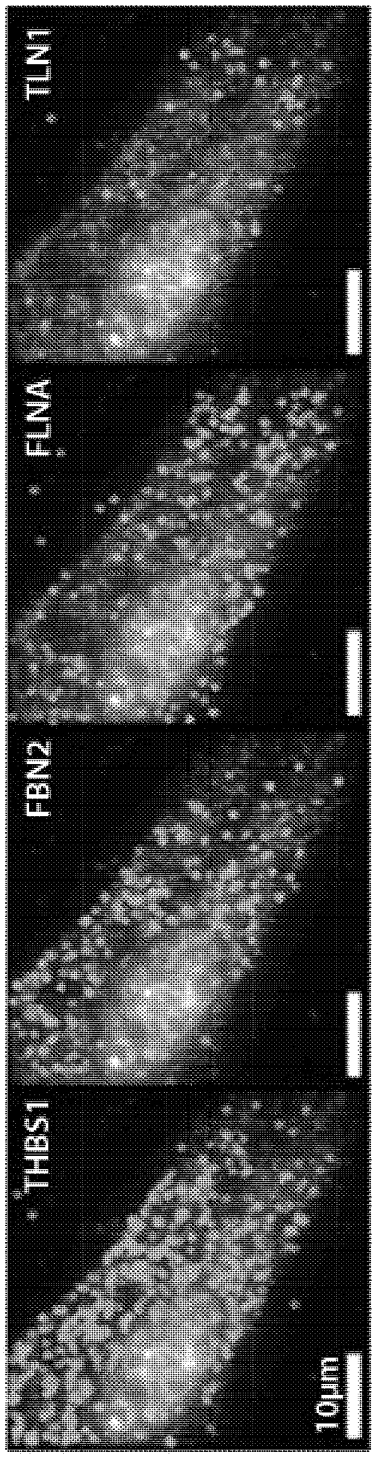


Fig. 8A

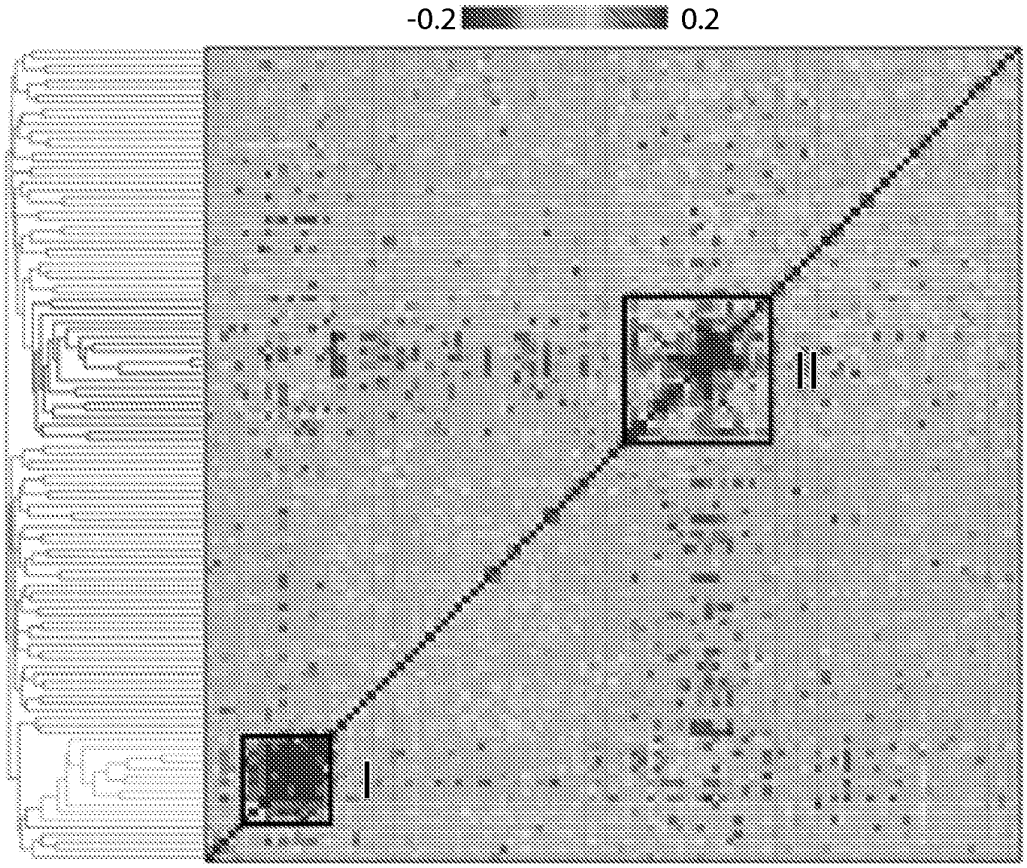


Fig. 8B

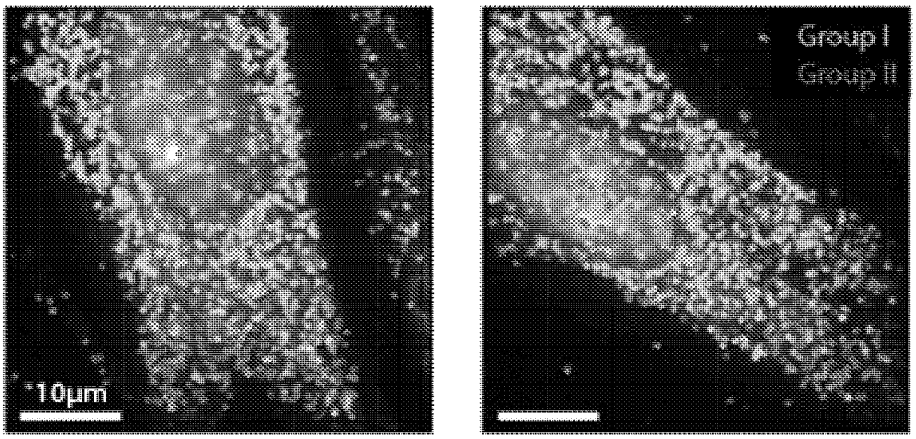


Fig. 8C

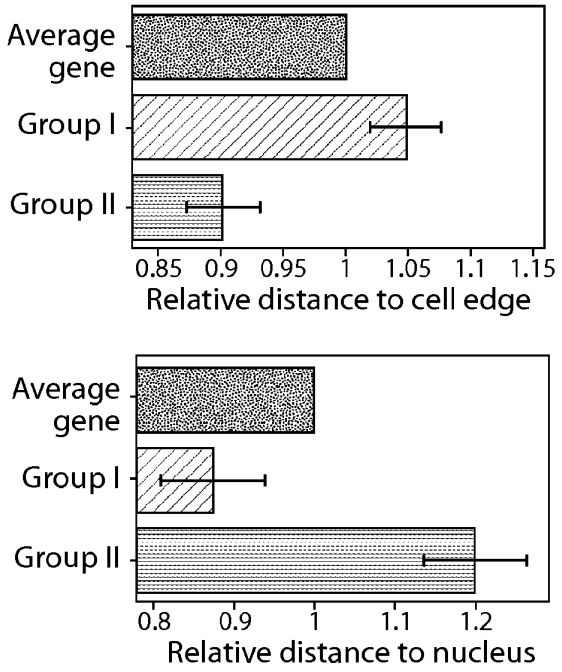

Average gene
Group I
Group II
0.85 0.9 0.95 1 1.05 1.1 1.15
Relative distance to cell edge
Average gene
Group I
Group II
0.8 0.9 1 1.1 1.2
Relative distance to nucleus

Fig. 8D

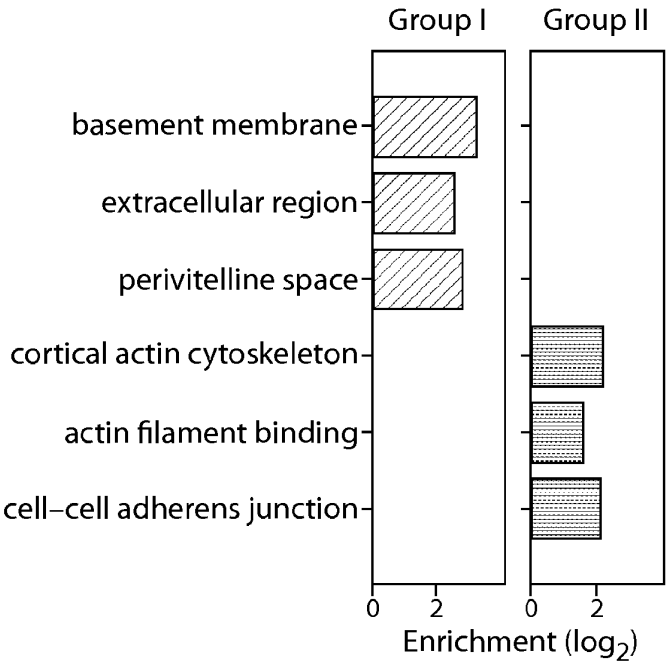

Group I
Group II
basement membrane
extracellular region
perivitelline space
cortical actin cytoskeleton
actin filament binding
cell–cell adherens junction
0 2
0 2
Enrichment (log2)

Fig. 8E

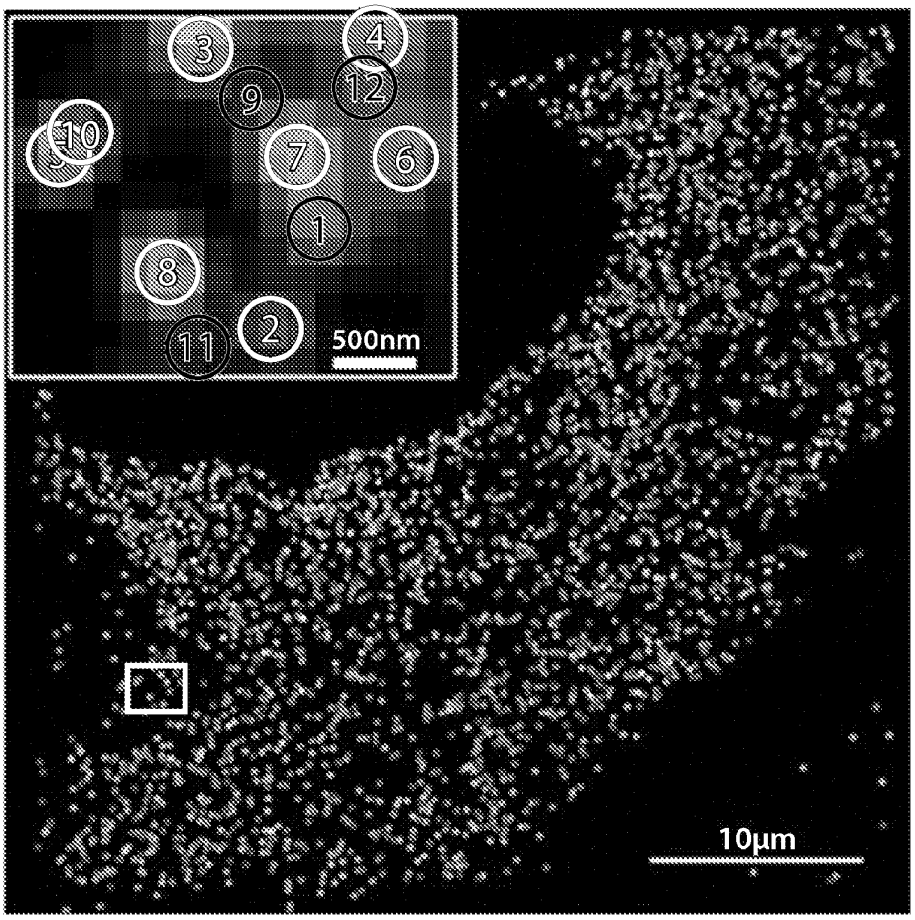


Fig. 9A

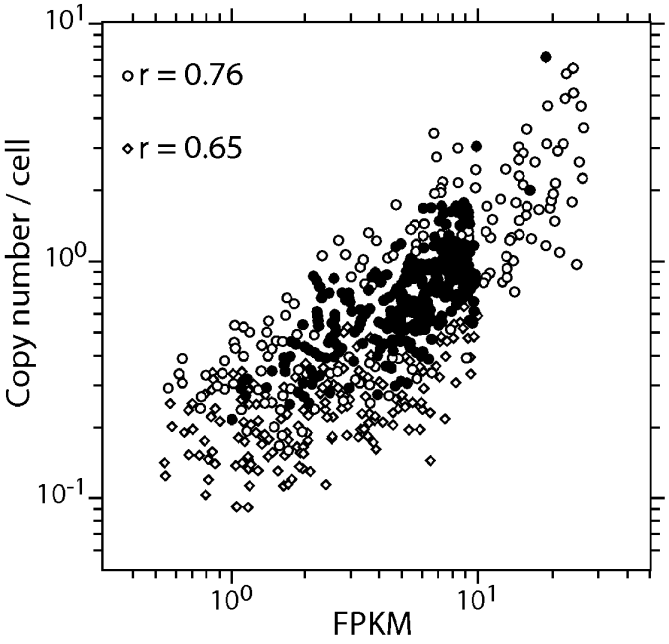

r = 0.76
r = 0.65
Copy number / cell
FPKM

Fig. 9B

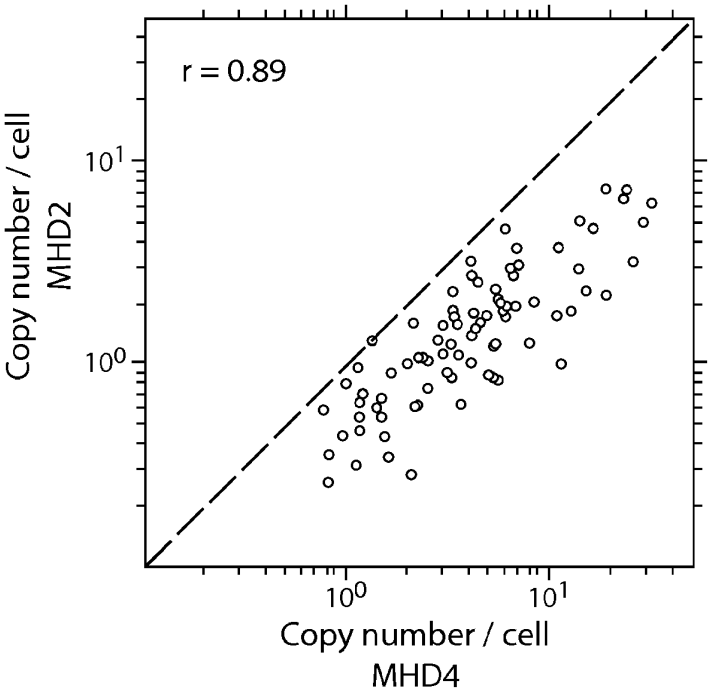

r = 0.89
Copy number / cell
MHD2
Copy number / cell
MHD4

Fig. 9C

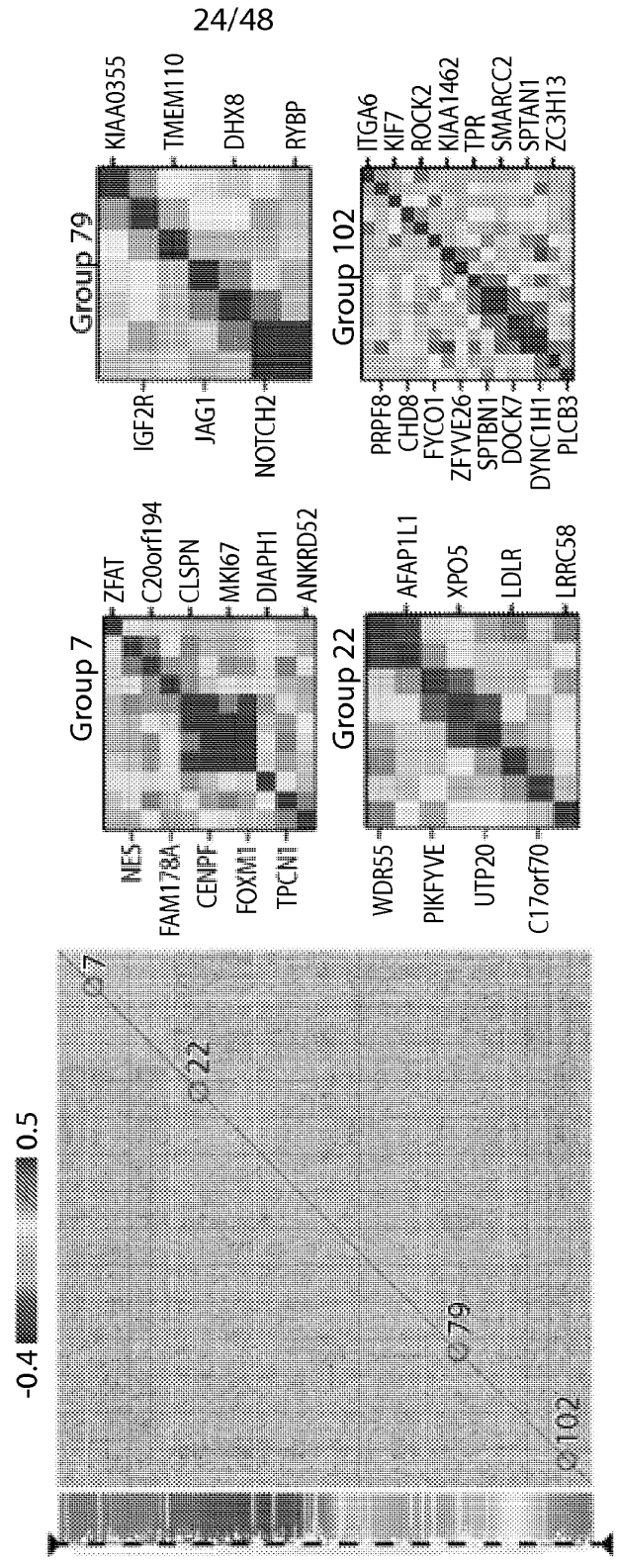

-0.4
0.5
g7
g22
g79
g102
Group 7
ZFAT
C20orf194
CLSPN
MKI67
DIAPH1
ANKRD52
NES
FAM178A
CENPF
FOXM1
TPCN1
Group 22
AFAP1L1
XPO5
LDLR
LRRC58
WDR55
PIKFYVE
UTP20
C17orf70
Group 79
KIAA0355
TMEM110
DHX8
RYBP
IGF2R
JAG1
NOTCH2
Group 102
ITGA6
KIF7
ROCK2
KIAA1462
TPR
SMARCC2
SPTAN1
ZC3H13
PRPF8
CHD8
FYCO1
ZFYVE26
SPTBN1
DOCK7
DYNC1H1
PLCB3

Fig. 10A

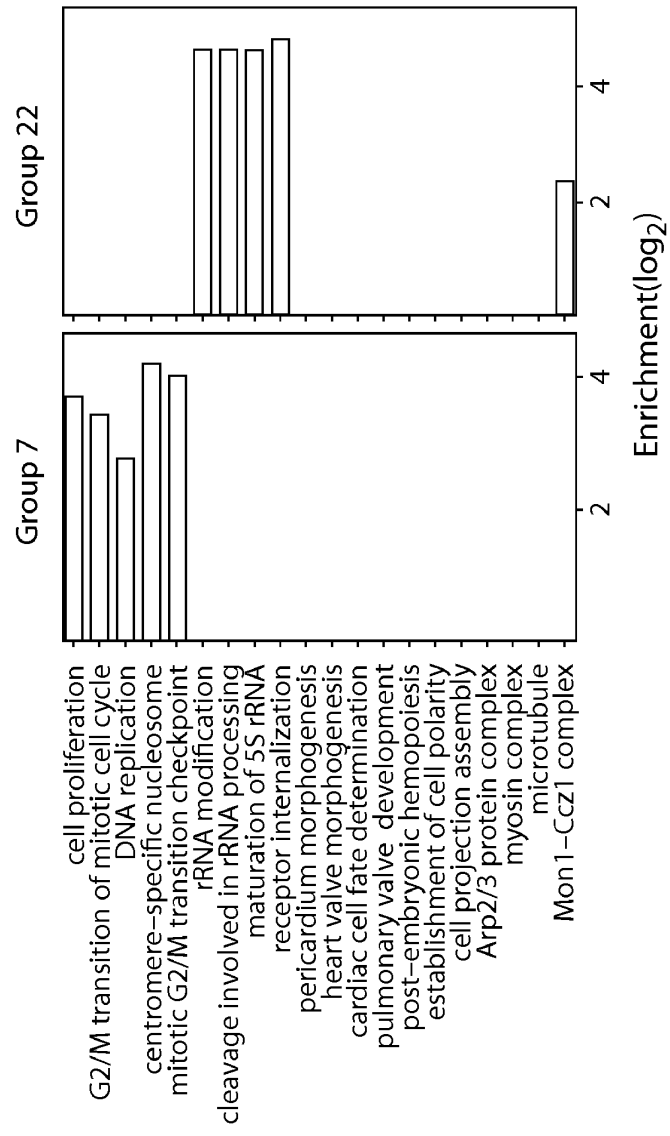

Group 7
Group 22
cell proliferation
G2/M transition of mitotic cell cycle
DNA replication
centromere−specific nucleosome
mitotic G2/M transition checkpoint
rRNA modification
cleavage involved in rRNA processing
maturation of 5S rRNA
receptor internalization
pericardium morphogenesis
heart valve morphogenesis
cardiac cell fate determination
pulmonary valve development
post−embryonic hemopoiesis
establishment of cell polarity
cell projection assembly
Arp2/3 protein complex
myosin complex
microtubule
Mon1−Ccz1 complex
2
4
2
4
Enrichment(log2)

Fig. 10B

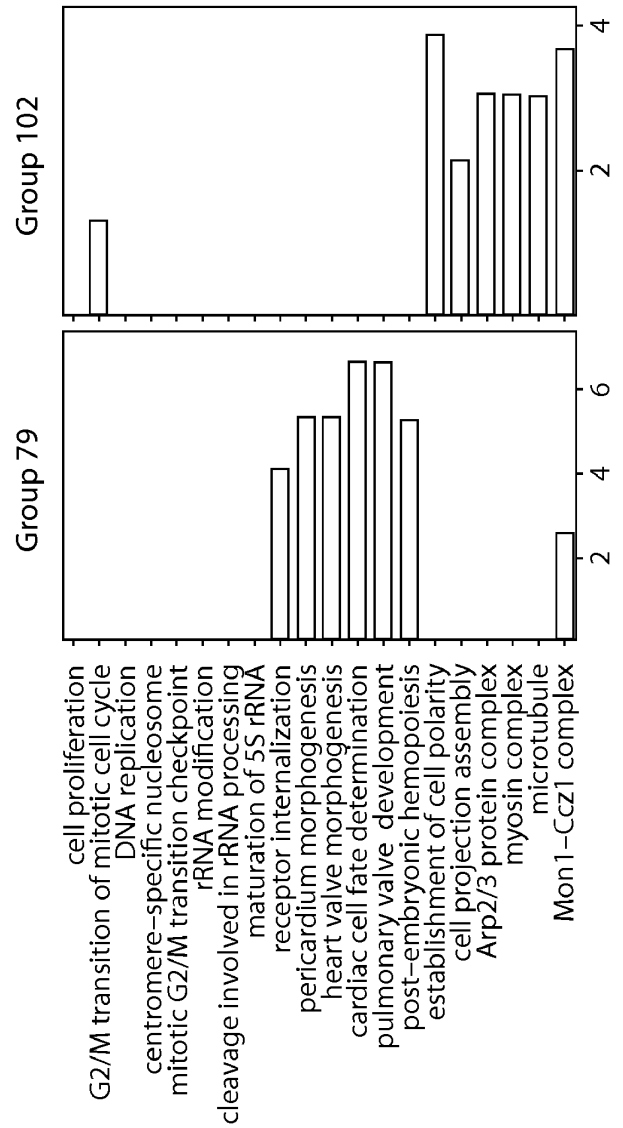


Fig. 10B Continued

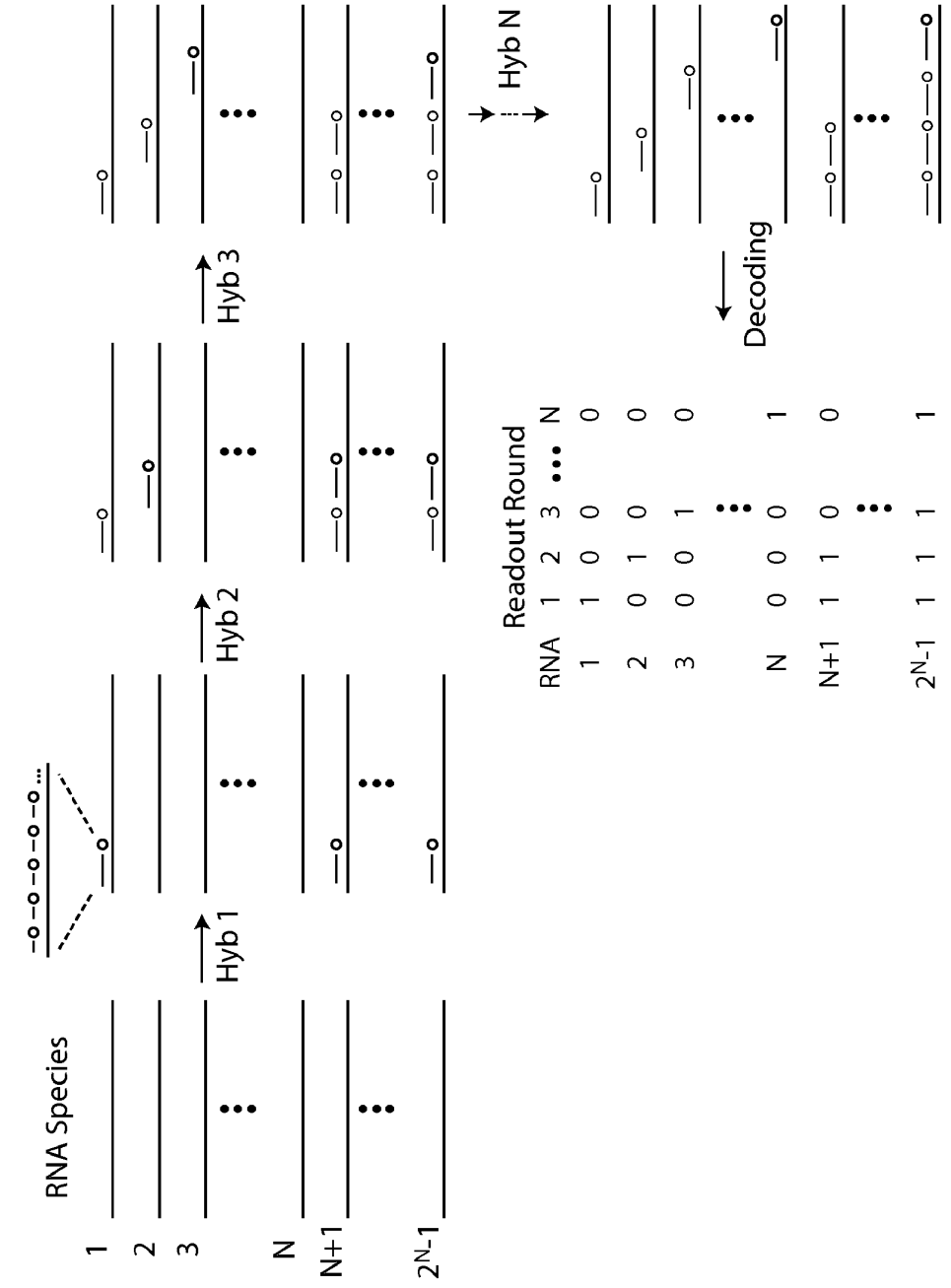

RNA Species
1
2
3
N
N+1
2N-1
Hyb 1
Hyb 2
Hyb 3
Hyb N
Readout Round
RNA 1 2 3 ... N
1 1 0 0 0
2 0 1 0 0
3 0 0 1 0
N 0 0 0 1
N+1 1 1 0 0
2N-1 1 1 1 1
Decoding

Fig. 11

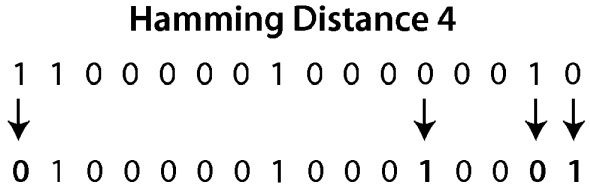


Fig. 12A

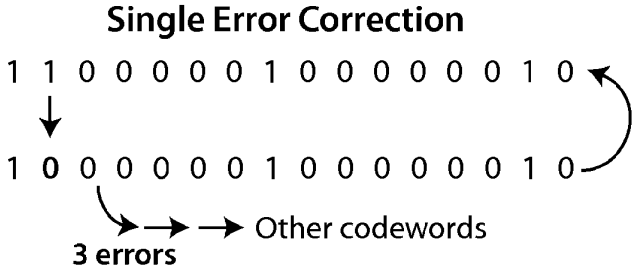


Fig. 12B

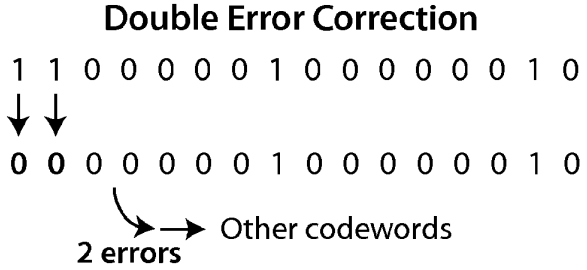


Fig. 12C

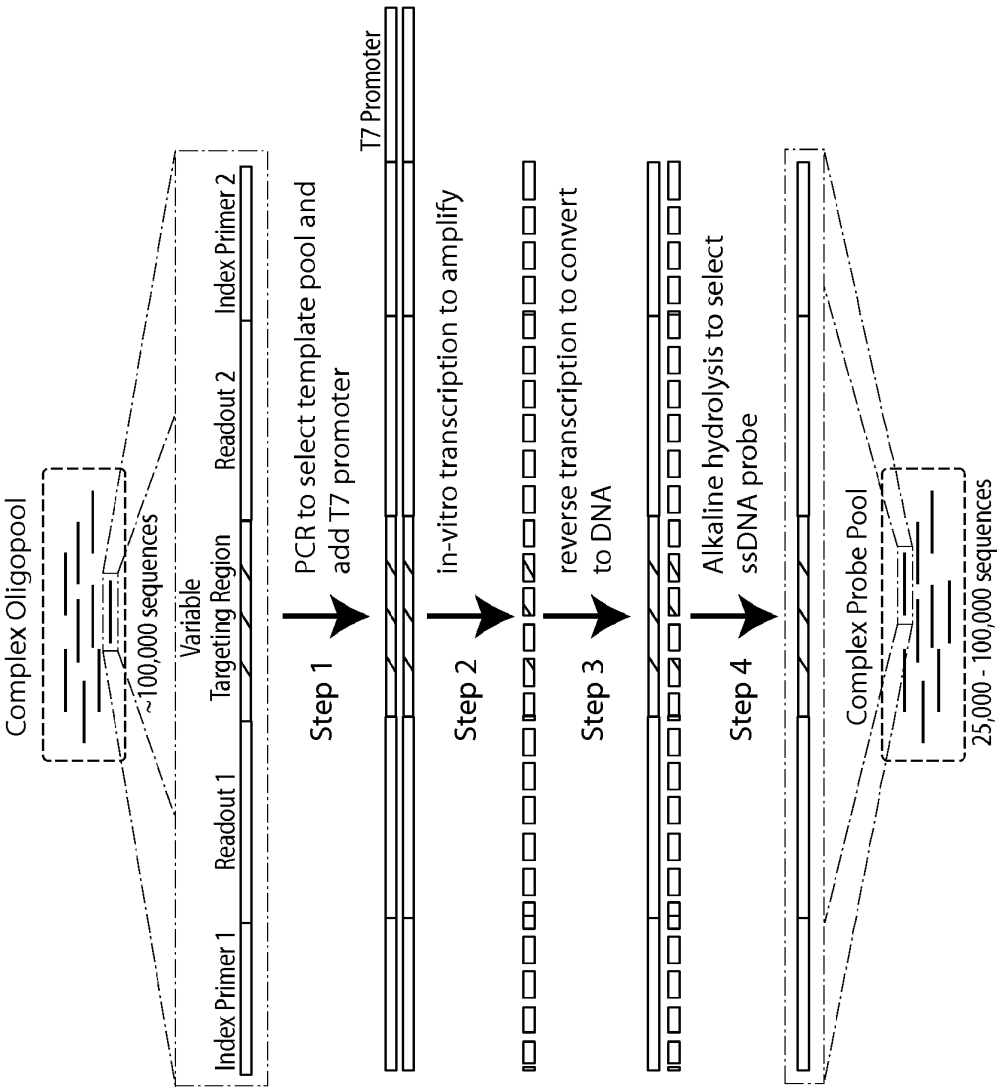

Complex Oligopool
~100,000 sequences
Index Primer 1
Readout 1
Variable Targeting Region
Readout 2
Index Primer 2
Step 1
PCR to select template pool and add T7 promoter
T7 Promoter
Step 2
in-vitro transcription to amplify
Step 3
reverse transcription to convert to DNA
Step 4
Alkaline hydrolysis to select ssDNA probe
Complex Probe Pool
25,000 - 100,000 sequences

Fig. 13

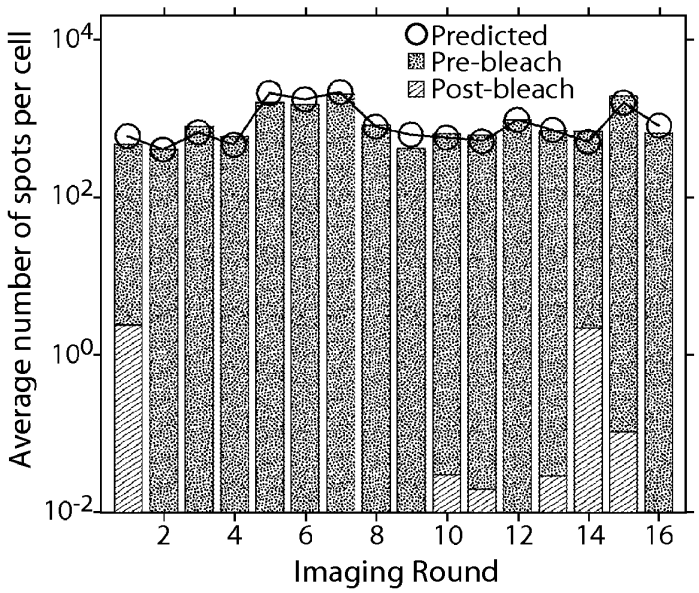


Fig. 14A

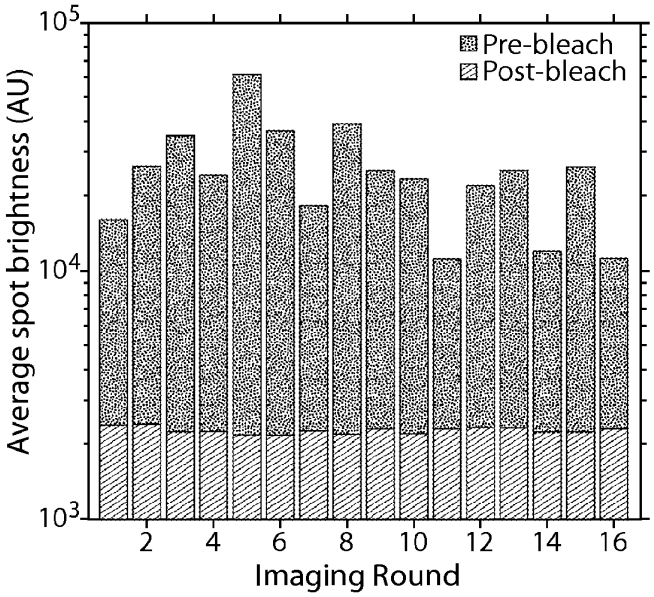


Fig. 14B

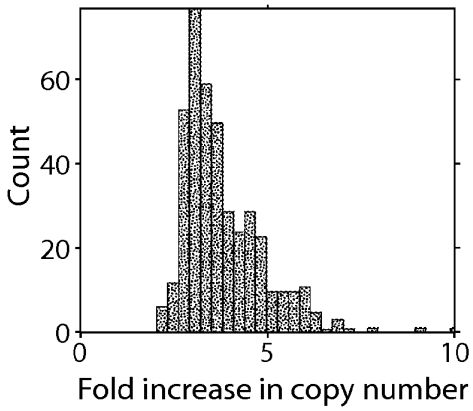


Fig. 15A

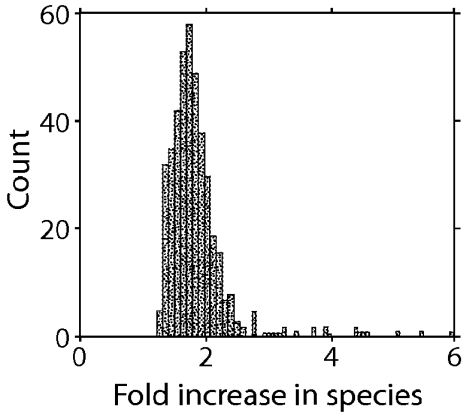


Fig. 15B

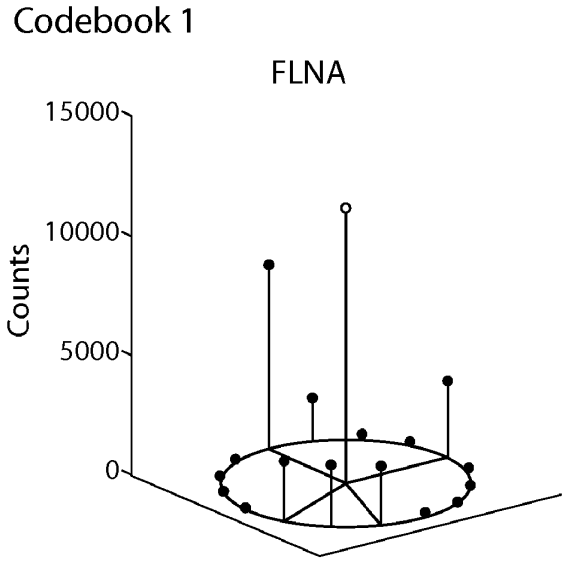


Fig. 16A

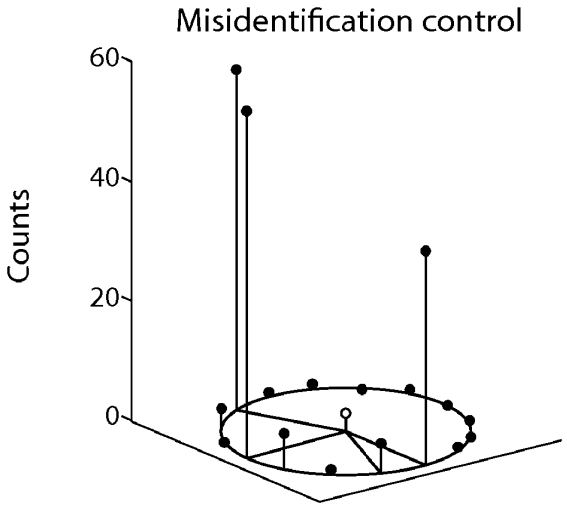


Fig. 16B

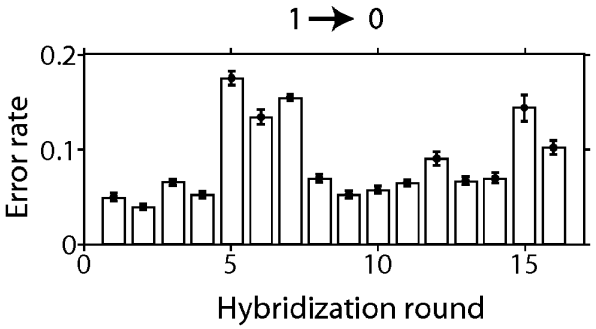


Fig. 16C

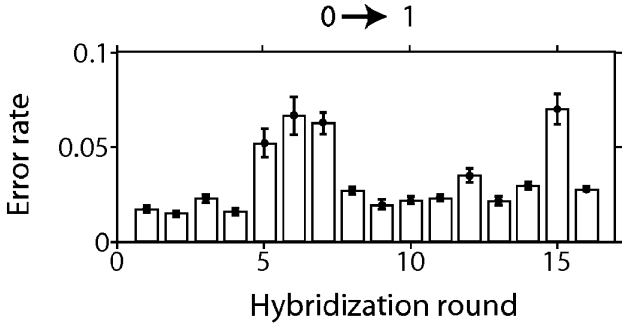


Fig. 16D

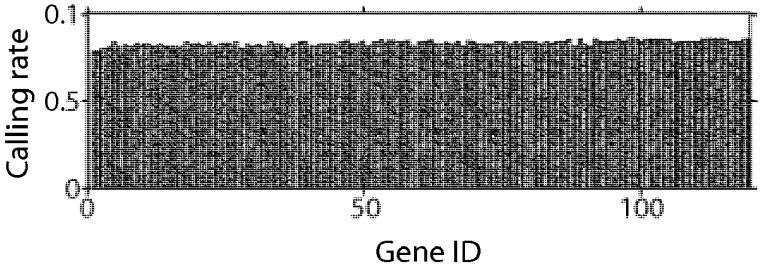


Fig. 16E

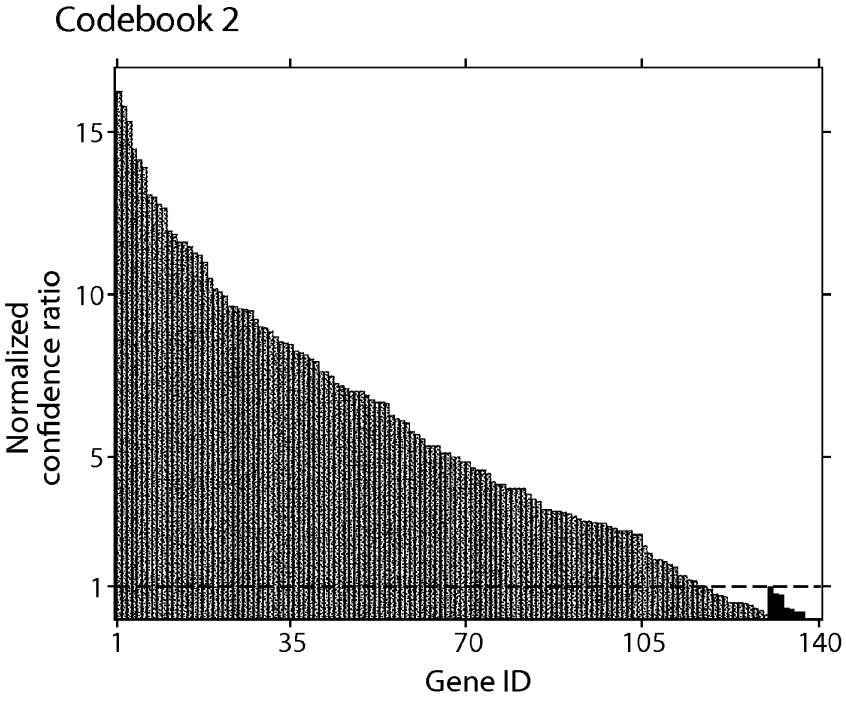


Fig. 17A

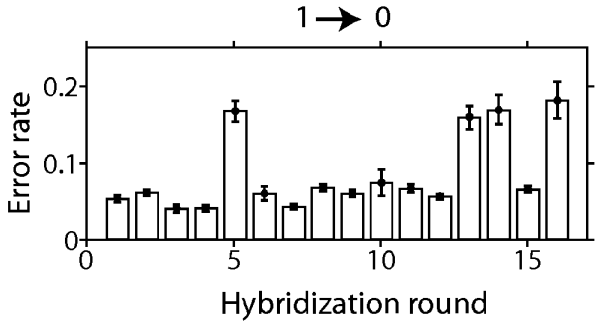


Fig. 17B

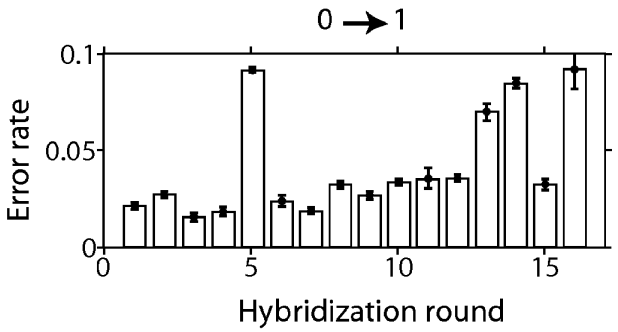


Fig. 17C

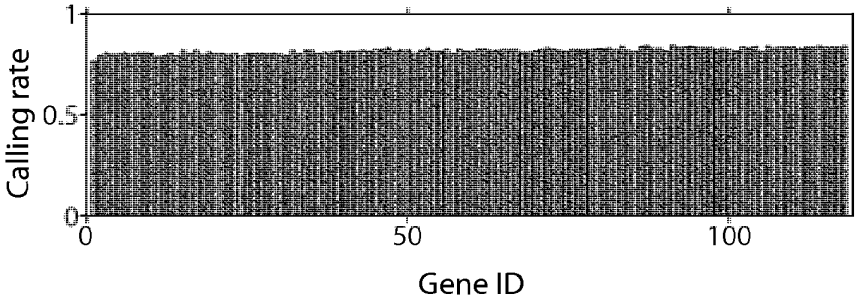


Fig. 17D

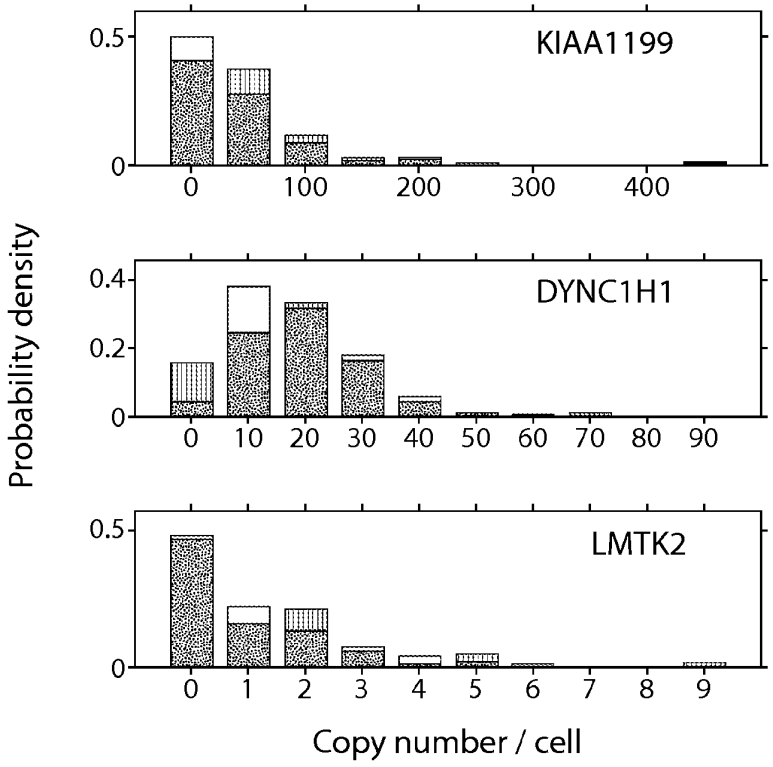

KIAA1199
DYNC1H1
LMTK2
Probability density
Copy number / cell
0.5
0
0.4
0.2
0 100 200 300 400
0 10 20 30 40 50 60 70 80 90
0 1 2 3 4 5 6 7 8 9

Fig. 18A

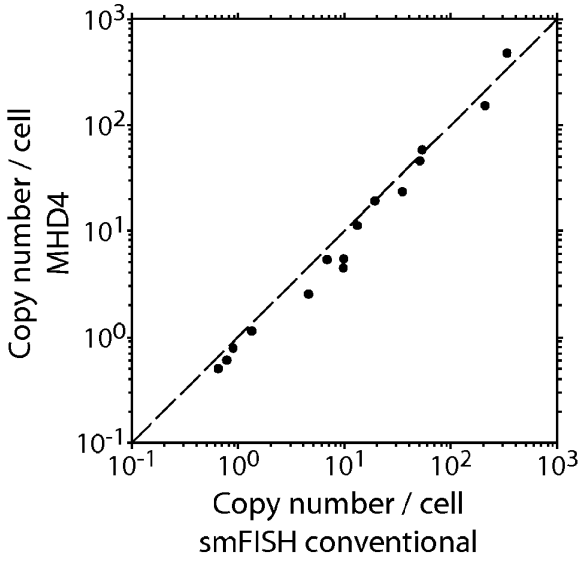


Fig. 18B

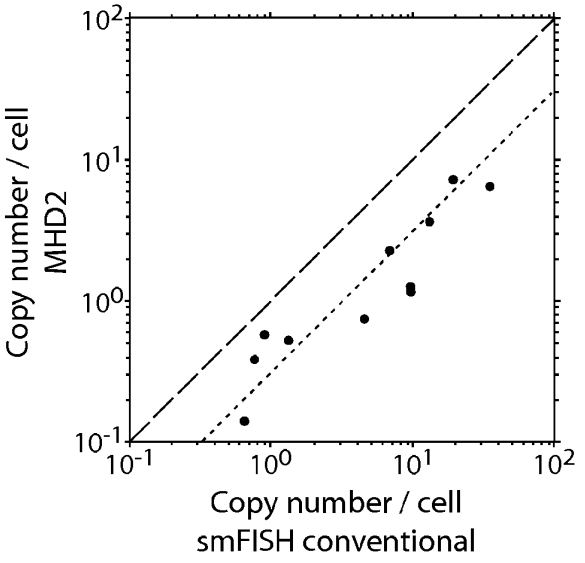


Fig. 18C

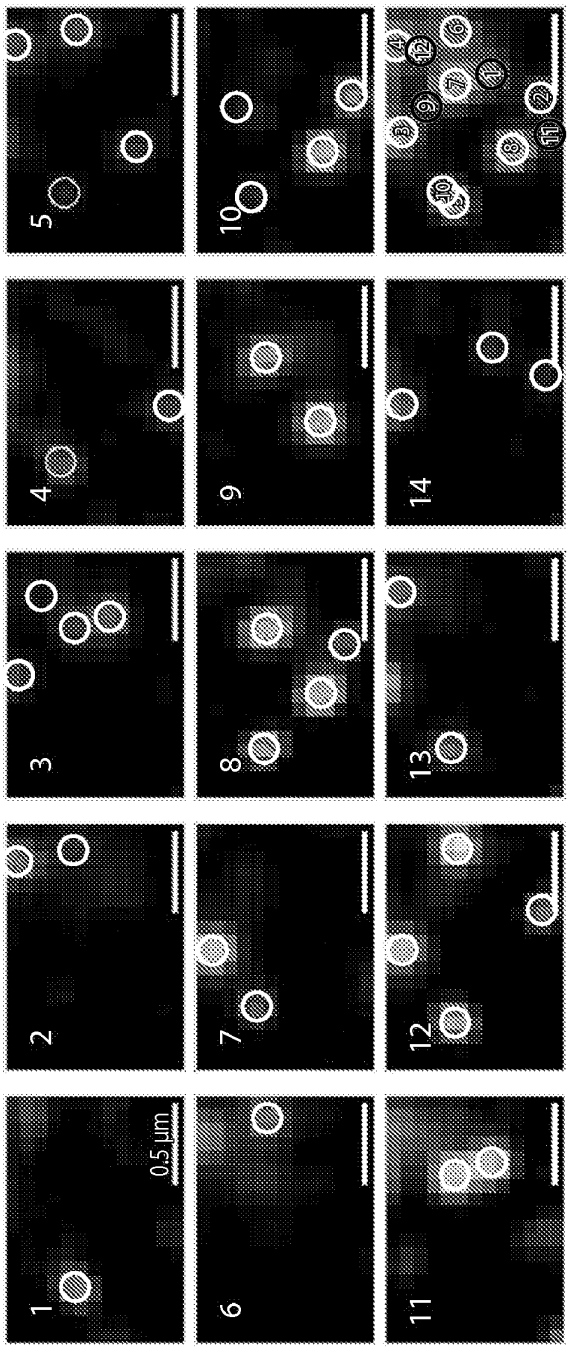


Fig. 19A

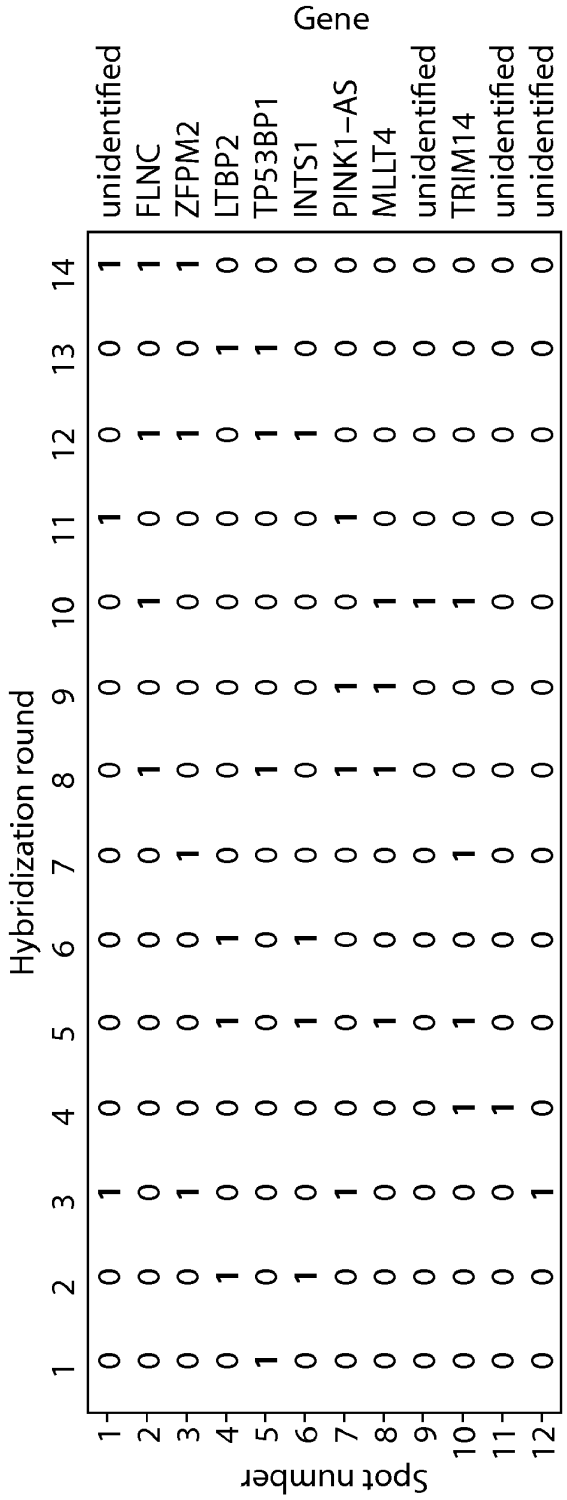

Gene
unidentified
FLNC
ZFPM2
LTBP2
TP53BP1
INTS1
PINK1-AS
MLLT4
unidentified
TRIM14
unidentified
unidentified
Hybridization round
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Spot number
1 0 0 1 0 0 0 0 0 0 1 0 0 1
2 0 0 0 0 0 0 0 1 0 1 0 1 0 1
3 0 0 1 0 0 0 1 0 0 0 0 1 0 1
4 0 1 0 0 1 0 0 0 0 0 0 0 1 0
5 1 0 0 0 0 0 0 1 0 0 0 1 1 0
6 0 1 0 0 1 1 0 0 0 0 0 1 0 0
7 0 0 1 0 0 0 0 1 1 0 1 0 0 0
8 0 0 0 0 1 0 0 1 1 1 0 0 0 0
9 0 0 0 0 0 0 0 0 0 1 0 0 0 0
10 0 0 0 1 1 0 1 0 0 1 0 0 0 0
11 0 0 0 1 0 0 0 0 0 0 0 0 0 0
12 0 0 1 0 0 0 0 0 0 0 0 0 0 0

Fig. 19B

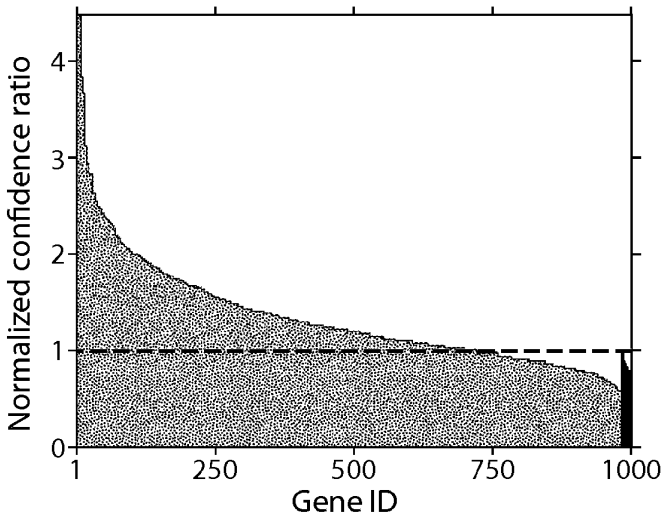


Fig. 19C

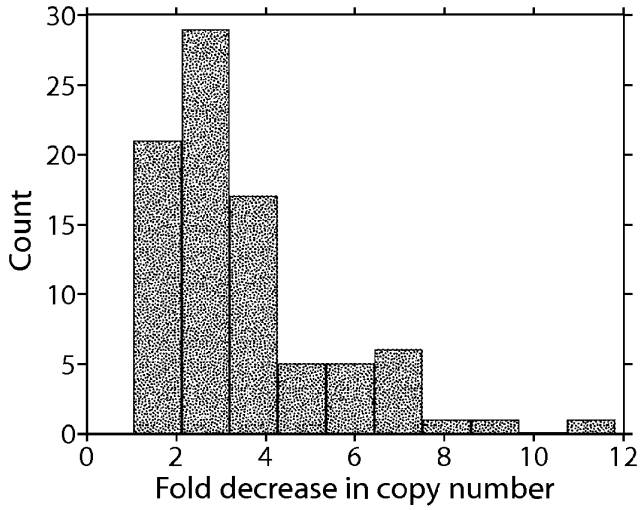


Fig. 19D

Codebook 1 for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
FLNA	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
TNC	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
FBN2	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
KIAA1199	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
COL5A1	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
MALAT1	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
TLN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FBN1	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
SRRM2	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
TEAD1	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
IQGAP1	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
MYH10	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
PAPPA	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
NR4A1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
LRP1	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
SLC38A1	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
SMARCA5	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
CKAP5	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
PIEZO1	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
VCAN	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
PRRC2B	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
SON	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
SLC7A11	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
GPR107	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
KIAA1462	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
notarget001	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
blank001	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0
ROCK2	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
SPTBN1	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FLNC	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
PRPF8	0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
IGF2R	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
FOSB	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
UBR5	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
SVEP1	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
CBX5	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
PDS5A	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
NUMA1	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
PLXNA1	0 0 0 1 0 0 0 0 0 0 0 1 1 0 1 0

Fig. 20-1

Codebook 1 for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
AFAP1	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
SPTAN1	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
DOCK7	0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0
NOTCH2	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
DYNC1H1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
ITGA2	1 0 1 0 0 0 0 0 0 0 0 0 1 1 0
PRKDC	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
GTF3C1	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
THSD4	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
PTPN14	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
LUZP1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
AMOTL1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
MAN1A2	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
notarget002	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
blank002	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
USP9X	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
CREBBP	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
KPNA4	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
EGFR	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
ZC3H13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
TNRC6A	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
TPR	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
COL7A1	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
ASCC3	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
USP34	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
CENPF	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
KIDINS220	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
SCUBE3	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
NRIP1	0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
NUP98	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
ZFC3H1	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
FAM208B	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
RAB3B	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
AFF4	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
PRKCA	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
FLNB	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
CHD8	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ANKRD52	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
MKI67	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0

Fig. 20-2

Codebook 1 for the 140-gene experiments

Genes	Codewords
DIP2C	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
SSH1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
notarget003	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
blank003	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
DNAJC13	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
DSEL	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PIK3CA	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
MED14	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
HERC2	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
H6PD	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
C14orf132	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
SIPA1L3	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
C17orf51	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
NF1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
VPS13A	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
ABCA2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0
PCNX	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
NKTR	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
FYCO1	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
ATP11A	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
REV3L	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
CASC5	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
AKAP11	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0
CEP250	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
USP24	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
CBL	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
RNF169	0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1
TMOD2	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FASN	0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
notarget004	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
blank004	0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 0
UBR2	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
PHIP	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
GTF3C4	0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
QSER1	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
VPS13D	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
FRY	0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 0 0
DIP2B	0 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0
HIVEP2	1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0

Fig. 20-3

Codebook 1 for the 140-gene experiments

Genes	Codewords
DOPEY1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
ALMS1	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
HEATR5B	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
HIVEP1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ARL10	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
HELZ2	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
POLQ	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
CACNA1A	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
KIAA2018	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
LYST	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
ZKSCAN2	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
DIEXF	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
FZD4	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
SLC5A3	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
KIAA1147	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0
LMTK2	0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0
STARD9	0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0
BRCA2	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
notarget005	1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20-4

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
FLNA	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
TNC	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
FBN2	0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1
KIAA1199	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
COL5A1	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
MALAT1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
TLN1	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FBN1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
SRRM2	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
TEAD1	1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0
IQGAP1	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
MYH10	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
PAPPA	0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0
NR4A1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
LRP1	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
SLC38A1	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
SMARCA5	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
CKAP5	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
PIEZO1	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
VCAN	0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0
PRRC2B	1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
SON	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
SLC7A11	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
GPR107	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
KIAA1462	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
notarget001	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
blank001	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
ROCK2	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
SPTBN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FLNC	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
PRPF8	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
IGF2R	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FOSB	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
UBR5	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
SVEP1	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
CBX5	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
PDS5A	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
NUMA1	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
PLXNA1	0 1 0 1 0 0 0 0 0 0 1 0 0 1 0 0

Fig. 20-5

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
ALPK2	0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0
AFAP1	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
SPTAN1	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
DOCK7	0 0 1 0 0 0 0 0 0 0 1 1 0 1 0 0
NOTCH2	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
DYNC1H1	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
ITGA2	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
PRKDC	0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
GTF3C1	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
THSD4	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
PTPN14	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
LUZP1	1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0
AMOTL1	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
MAN1A2	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
notarget002	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
blank002	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
USP9X	0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
CREBBP	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1
KPNA4	1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1
EGFR	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1
ZC3H13	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
TNRC6A	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
TPR	1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0
COL7A1	0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1
ASCC3	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
USP34	0 0 1 0 0 0 1 0 0 0 1 0 1 0 0 0
CENPF	0 0 1 0 0 0 0 0 1 0 0 1 1 0 0 0
KIDINS220	0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 1
SCUBE3	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
NRIP1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
NUP98	0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0
ZFC3H1	0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1
FAM208B	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
RAB3B	0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0
AFF4	0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1
PRKCA	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
FLNB	0 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0
CHD8	0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
ANKRD52	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
MKI67	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0

Fig. 20-6

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DIP2C	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
SSH1	0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0
notarget003	0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1
blank003	0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0
DNAJC13	0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0
DSEL	0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
PIK3CA	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
MED14	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
HERC2	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
H6PD	0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0
C14orf132	0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0
SIPA1L3	1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
C17orf51	0 1 0 0 0 0 0 0 0 1 1 0 1 0 0 0
NF1	1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0
VPS13A	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
ABCA2	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
PCNX	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
NKTR	0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0
FYCO1	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
ATP11A	1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0
REV3L	1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0
CASC5	0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0
AKAP11	0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 0
CEP250	1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0
USP24	1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0
AGPS	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
CBL	0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0
RNF169	0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
TMOD2	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
FASN	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
notarget004	0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
blank004	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
UBR2	0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0
PHIP	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
GTF3C4	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
QSER1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
VPS13D	1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
FRY	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
DIP2B	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
HIVEP2	1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0

Fig. 20-7

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
DOPEY1	0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 0
ALMS1	0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0
HEATR5B	0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0
HIVEP1	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
ARL10	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
HELZ2	0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0
POLQ	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
CACNA1A	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
KIAA2018	0 0 0 0 0 0 0 1 1 0 1 0 0 0 1 0
LYST	0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0
ZKSCAN2	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
DIEXF	0 1 0 0 0 0 0 1 0 0 1 1 0 0 0 0
FZD4	0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0
SLC5A3	1 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0
KIAA1147	0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0
LMTK2	0 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0
STARD9	0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0
BRCA2	1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0
notarget005	1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
blank005	1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0

Fig. 20-8

721258

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2015

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *62/050,636*
FILING DATE: *September 15, 2014*
RELATED PCT APPLICATION NUMBER: *PCT/US15/42556*

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS *US62/050,636*

Certified by

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PTO/SB/16 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET – Page 1 of 2

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. Via EFS-Web

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Xiaowei	Zhuang	Lexington, Massachusetts
Jeffrey R.	Moffitt	Somerville, Massachusetts
Alistair	Boettiger	Cambridge, Massachusetts

Additional inventors are being named on the _____ separately numbered sheets attached hereto.

TITLE OF THE INVENTION (500 characters max):

PROBE LIBRARY CONSTRUCTION

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

[X] The address corresponding to Customer Number: 86110

OR

[ ] Firm or Individual Name

Address

City	State	Zip
Country	Telephone	Email

ENCLOSED APPLICATION PARTS (*check all that apply*)

[X] Application Data Sheet. See 37 CFR 1.76.
[ ] CD(s), Number of CDs _____
[X] Drawing(s) *Number of Sheets* 39
[X] Other (specify) Appendix A – 77 pages
[X] Specification (e.g. description of the invention) *Number of Pages* 48

Fees Due: Filing Fee of $260 ($130 for small entity) ($65 for micro entity). If the specification and drawings exceed 100 sheets of paper, an application size fee is also due, which is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

METHOD OF PAYMENT OF THE FILING FEE AND APPLICATION SIZE FEE FOR THIS PROVISIONAL APPLICATION FOR PATENT

[X] Applicant asserts small entity status. See 37 CFR 1.27.
[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.
[ ] A check or money order made payable to the *Director of the United States Patent and Trademark Office* is enclosed to cover the filing fee and application size fee (if applicable).
[ ] Payment by credit card. Form PTO-2038 is attached.

130.00
TOTAL FEE AMOUNT ($)

[X] The Director is hereby authorized to charge the filing fee and application size fee (if applicable) or credit any overpayment to Deposit Account Number: 23/2825 .

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

3728774.1

PTO/SB/16 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION COVER SHEET - Page 2 of 2

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[ ] No.

[ ] Yes, the invention was made by an agency of the United States Government. The U.S. Government agency name is:

[X] Yes, the invention was made under a contract with an agency of the U.S. Government. The name of the U.S. Government agency and the Government contract number are:

National Institutes of Health - GM 096450

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

SIGNATURE	/Tani Chen/	DATE	September 15, 2014
TYPED or PRINTED NAME	Tani Chen, Sc.D.	REGISTRATION NO. *(if appropriate)*	52,728
TELEPHONE	617.646.8000	DOCKET NUMBER	H0498.70510US00

Electronic Acknowledgement Receipt	
EFS ID:	20144276
Application Number:	62050636
International Application Number:	
Confirmation Number:	1939
Title of Invention:	PROBE LIBRARY CONSTRUCTION
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Angela Griffith
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70510US00
Receipt Date:	15-SEP-2014
Filing Date:	
Time Stamp:	18:37:30
Application Type:	Provisional

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$130
RAM confirmation Number	5324
Deposit Account	232825
Authorized User	CHEN, TANI
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Provisional Cover Sheet (SB16)	H049870510US00-PROCVR-TC.pdf	39019 a5a34c3fc1348c18f1c4e76f4f866eca052851b4	no	2
Warnings:					
This is not a USPTO supplied Provisional Cover Sheet SB16 form.					
Information:					
2	Fee Worksheet (SB06)	H049870510US00-FEE-TC.pdf	30341 cd4c142d2a4dc0ea029da6723c7c7a3c6a17fe49	no	1
Warnings:					
Information:					
3	Application Data Sheet	H049870510US00-ADS-TC.pdf	20989 d49b53eae6f1df3793273acdfdaa66441509eea3	no	4
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
4		H049870510US00-PROAPL-TC.pdf	194507 fbccdc5ab7322a0db9c186b3ba0490738fc9e310	yes	48

Multipart Description/PDF files in .zip description

Document Description	Start	End
Specification	1	31
Claims	32	47
Abstract	48	48

Warnings:					
Information:					
5	Drawings-other than black and white line drawings	H049870510US00-DRW-TC.pdf	135443 d63a5a27c564d6a0d669fdb63de6d6d423e26220	no	39
Warnings:					
Information:					
6	Appendix to the Specification	H049870510US00-APDX-TC.pdf	738932 bd297c5326ac22a91e6514aecc1dbd9a54b558be	no	77
Warnings:					
Information:					

7	Fee Worksheet (SB06)	fee-info.pdf	29901 4b989fa2c3282cff3ef7255c81d4119751dbf747	no	2

Warnings:	
Information:	
Total Files Size (in bytes):	1189132

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Xiaowei
Family Name::	Zhuang
City of Residence::	Lexington
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	5 Luongo Farm Lane
City of mailing address::	Lexington
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02421

Inventor Number::	2
Given Name::	Jeffrey
Middle Name::	R.
Family Name::	Moffitt
City of Residence::	Somerville
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	175 Beacon Street
	Apt. #108
City of mailing address::	Somerville

State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02143

Inventor Number::	3
Given Name::	Alistair
Family Name::	Boettiger
City of Residence::	Somerville
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	1550 Cambridge St, Apt 2
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02139

Correspondence Information

Correspondence Customer Number::	86110

Application Information

Application Type::	Provisional
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No

Title::	PROBE LIBRARY CONSTRUCTION
Attorney Docket Number::	H0498.70510US00
Request for Early Publication?::	No
Request for Non-Publication?::	No
Small Entity?::	Yes
Petition included?::	No
Licensed US Govt. Agency::	National Institutes of Health
Contract or Grant Numbers::	GM 096450

Representative Information

Representative Customer Number::	86110

Domestic Priority Information

Foreign Priority Information

Applicant Information

Applicant Number::	1
Applicant Type::	Assignee
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge

State or Province of mailing address:: MA

Postal or Zip Code of mailing address:: 02138

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.			
Signature	/Tani Chen/	Date	September 15, 2014
Name (Print/Type)	Tani Chen, Sc.D.	Registration No. (Attorney/Agent)	52,728

PROBE LIBRARY CONSTRUCTION

GOVERNMENT FUNDING

Research leading to various aspects of the present invention was sponsored, at least in part, by the NIH, Grant No. GM096450. The U.S. Government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for producing nucleic acids.

BACKGROUND

Custom-synthesized, oligonucleotide probes have emerged as a powerful tool for the identification and isolation of specific nucleic acid targets via hybridization. Applications for such hybridization probe sets range from next generation sequencing—where such probes are used to enrich or deplete samples for specific nucleic acid targets—to imaging of fixed samples—where fluorescently labeled hybridization probes allow the direct measurement of the number and spatial organization of the targeted species.

There are now a wide range of commercial sources for such probes. Such probes are often made by synthesizing each oligonucleotide member using standard solid phase synthesis methods. Unfortunately, this limits both the number of probes within a single set and the number of unique sets, due to the requirement that each oligonucleotide member must be individually and separately synthesized.

Recent advances in array-based synthesis of oligonucleotides by several companies have reduced the cost of producing oligonucleotides. However, these approaches also result in 1000-fold less oligonucleotide probes than is required for a single hybridization reaction, thus limiting their usefulness. Accordingly, improvements in oligonucleotide production are needed.

SUMMARY

The present invention generally relates to systems and methods for producing nucleic acids. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a method. The method, in accordance with one set of embodiments, includes amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides, transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA, reverse transcribing the RNA to produce transcribed DNA, and selectively degrading the RNA relative to the transcribed DNA.

In another set of embodiments, the method includes simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides, transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA, reverse transcribing the RNA to produce transcribed DNA, and selectively degrading the RNA relative to the transcribed DNA.

In yet another set of embodiments, the method includes acts of providing a plurality of oligonucleotides having an average length of between 10 and 200 nucleotides and including at least 100 unique oligonucleotide sequences, producing amplified oligonucleotides comprising one of the plurality of oligonucleotides and a promoter using real-time PCR, transcribing at least some of the amplified oligonucleotides to produce RNA using an RNA polymerase, reverse transcribing the RNA to produce DNA using a primer comprising a signaling entity, and chemically reducing the RNA.

The method, in still another set of embodiments, includes acts of providing a plurality of oligonucleotides having an average length of between 10 and 200 nucleotides and including at least 100 unique oligonucleotide sequences, producing amplified oligonucleotides in a common solution comprising one of the plurality of oligonucleotides and a promoter using PCR, transcribing at least some of the amplified oligonucleotides to produce RNA using an RNA polymerase, reverse transcribing the RNA to produce DNA using a primer comprising a signaling entity, and chemically reducing the RNA.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein, such as oligonucleotides, including but not limited to modified oligonucleotides such as those described herein (e.g., labeled with a signaling entity). In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein, for example, such as

oligonucleotides, including but not limited to modified oligonucleotides such as those described herein (e.g., labeled with a signaling entity).

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Fig. 1 illustrate the production of DNA probes, in accordance with one set of embodiments; and

Fig. 2 illustrates a template molecule produced in accordance with one embodiment of the invention;

Fig. 3 lists optimized primers from the *E. coli* transcriptome, in another embodiment of the invention; and

Fig. 4 shows various probes in accordance with yet another embodiment of the invention.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for producing nucleic acids. In some aspects, relatively large quantities of oligonucleotides can be produced, and in some cases, the oligonucleotides may have a variety of different sequences and/or lengths. For instance, a relatively small quantity of oligonucleotides may be amplified to produce a large amount of nucleotides. In one set of embodiments,

oligonucleotides may be amplified using PCR, then transcribed to produce RNA. The RNA may then be reverse transcribed to produce DNA, and optionally, the RNA may be selectively degraded or removed, relative to the DNA. In one set of embodiments, the oligonucleotides may be chemically modified. These modifications may include, but are not limited, to the adding of fluorescent dyes or other signaling entities.

U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.* is incorporated herein by reference in its entirety. A copy is provided herein as Appendix A.

In one aspect, the present invention is generally directed to *in vitro* methods of amplifying a plurality of oligonucleotides. In some cases, relatively large numbers of unique oligonucelotides within a plurality of oligonucleotides may be amplified. For instance, a plurality of oligonucleotides to be amplified may include 10, 100, 1,000, or more unique sequences.

In addition, in some embodiments, the oligonucleotides may be amplified without selective amplification of some oligonucleotides over others, e.g., due to competitive effects. Although some drift may occur, it is desired that the relative ratios of the oligonucleotides within a plurality of oligonucleotides stay substantially the same after amplification, at least for some applications. However, in many amplification techniques, due to differences in binding or affinity of different oligonucleotides, some oligonucleotides may be amplified to a greater degree than others, and thus, specific techniques need to be utilized to reduce or eliminate this problem, for example, by separately amplifying each of the oligonucleotides before combining them together to form the plurality of oligonucleotides. In contrast, as is discussed herein, in certain embodiments, a plurality of oligonucleotides can be amplified without causing substantial alterations or changes in the ratios of the oligonucleotides, without requiring separation of the oligonucleotides, separate growth of the oligonucleotides, or other cumbersome techniques.

Referring now to Fig. 1, one example of an embodiment of the invention is now illustrated. In this figure, a plurality of oligonucleotides 10 is provided. This may include 1, 10, 100, 1,000, 10,000, 100,000, or any other suitable number of unique oligonucleotide sequences. Of course, more than one copy of any particular unique

oligonucleotide may also be present as well within the plurality of oligonucleotides. The unique oligonucleotides may have the same or different lengths. In some cases, the plurality of oligonucleotides have an overall average length (number average or arithmetic mean) of less than 200 nt (nucleotides), although longer average lengths are also possible in some embodiments.

The plurality of oligonucleotides 10 may initially be amplified, using PCR (polymerase chain reaction) or another suitable oligonucleotide amplification method, to produce a plurality of amplified oligonucleotides 20. In some cases, PCR may be used to generate thousands to millions of copies per oligonucleotide within the plurality of oligonucleotides. In some embodiments, the plurality of oligonucleotides may be amplified while still contained in a common solution, for instance, without requiring separation of the oligonucleotides prior to amplification, e.g., as is required in certain techniques such as emulsion PCR.

Within a common solution, while it is possible that different oligonucleotides of the plurality of oligonucleotides may be amplified at different rates (e.g., leading to non-uniform amplification of the plurality of oligonucleotides, and the potential loss of complexity or species within the plurality of oligonucleotides during amplification), in certain embodiments, this can be reduced or minimized through the use of various oligonucleotide structures and/or through the use of certain types of PCR techniques, as is discussed herein.

As an example, in one set of embodiments, the plurality of oligonucleotides may all be chosen to minimize competitive effects, e.g., as caused by differences in binding or affinity of the oligonucleotides to reagents within the common solution or the preferential enzymatic amplification of some sequence features. For example, in one set of embodiments, the plurality of oligonucleotides may be chosen to have similar lengths and/or sequences.

As is shown in Fig. 1, the plurality of oligonucleotides may each contain one or more index regions 15, 16 on one or both ends of the oligonucleotides that can be recognized by certain reagents. In some cases, the oligonucleotides of the plurality of oligonucleotides may have one or more index regions to which suitable primers can interact with in order to allow PCR or other amplification to occur. In some

embodiments, these index regions can be used to selectively produce DNA probes only from subset of the plurality of oligonucleotides 10.

These index regions can also be used in some instances to add additional sequences to that of the plurality of oligonucleotides, e.g., as is shown in Fig. 1, oligonucleotide 11 may include an index region 15, to which a sequence 17 containing a T7 promoter can bind and be introduced into the amplified oligonucleotides (e.g., as region 23). Various sequences may thus be applied to the plurality of oligonucleotides that include a portion able to bind an index region. In some cases, if substantially all of the plurality of oligonucleotides contain similar or identical index regions, the relative affinities or binding to the index regions of the oligonucleotides by enzymes such as polymerases may be substantially similar or identical, which may allow for relatively uniform amplification to occur. The plurality of oligonucleotides may also contain other different regions that can be varied to produce a plurality of unique oligonucleotides, e.g., region 12. These regions can vary in terms of length and/or sequence, etc.

In some embodiments, the amount of amplification that occurs may be carefully controlled by monitoring the PCR amplification reaction, e.g., using techniques such as real-time PCR. This may occur, for example, using oligonucleotides having common index regions, substantially similar lengths and/or sequences, etc., including those previously discussed, or with other suitable pluralities of oligonucleotides. For instance, in some embodiments, the PCR reaction may be monitored by illuminating the solution containing the oligonucleotides with suitable light and determining the amount of fluorescence that is present, which can be related to the DNA present within the sample. Techniques for monitoring PCR reactions, such as real-time PCR methodologies, are known to those of ordinary skill in the art. The PCR reaction can also be controlled, in some embodiments, by controlling the amount and/or concentration of nucleotides and/or cofactors that are present.

After amplification, the plurality of amplified oligonucleotides 20 may be transcribed to produce a plurality of RNAs 30, as is shown in the example of Fig. 1. This may be performed, for example, by exposing the amplified oligonucleotides to a suitable RNA polymerase, such as T7 RNA polymerase, that can transcribe the oligonucleotides to produce corresponding RNA. The amount of RNA production can be controlled in

some embodiments by controlling the amount and/or concentration of nucleotides and/or cofactors that are present as well as the duration of the *in vitro* transcription reaction.

The plurality of RNAs 30, may then be used to produce additional amounts of DNA 40, e.g., by using reverse transcription. For example a suitable enzyme, such as reverse transcriptase, may be used to perform the reverse transcription. In some cases, primers may be used to facilitate transcription, and in some embodiments, the primers may also be used to attach additional entities to the DNA. For example, signaling entities may be attached, as is shown with signaling entity 48 in Fig. 1. Alternatively, additional nucleic acid sequences can also be attached, which can serve to recruit additional oligonucleotides via Watson-Crick base-pairing. The amount of DNA that is produced can be controlled, for example, by controlling the amount and/or concentration of nucleotides and/or cofactors that are present as well as the duration and temperature of the reverse transcription reaction.

In some embodiments, multiple copies of DNA may be produced from each RNA molecule. In addition, optionally, the RNA may then be removed or selectively degraded, relative to the DNA, for example, through alkaline hydrolysis, enzymatic digestion, or other techniques.

Accordingly, in certain aspects, the present invention is generally directed to systems and methods of amplifying a plurality of oligonucleotides. In one set of embodiments, relatively large quantities or masses of oligonucleotides can be produced as is discussed herein, e.g., at least about 10^{-3} pmol, at least about 10^{-2} pmol, at least about 10^{-1} pmol, at least about 10^{0} pmol, at least about 10^{1} pmol, at least about 10^{2} pmol, at least about 10^{3} pmol, etc. In addition, in some embodiments, the plurality of oligonucleotides may be substantially diverse. For example, the plurality of oligonucleotides may include at least about 10^{1}, at least about 10^{2}, at least about 10^{3}, at least about 10^{4}, at least about 10^{5}, or at least about 10^{6} unique sequences of oligonucleotides, even after amplification to the amounts discussed above. (It should also be noted that a plurality or population of oligonucleotides may include more than one copy of a given unique oligonucleotide sequence.) In contrast, certain prior art techniques are able to amplify large numbers of unique oligonucleotides, but to only small quantities or masses (e.g., to amounts of around 10^{-3} pmol or less), or are able to

produce large quantities or masses of oligonucleotides, but only for 1 or a few unique sequences (e.g., less than 10 sequences).

As discussed, in certain embodiments, a plurality of oligonucleotides, which may include a plurality of unique sequences of oligonucleotides such as those described above, may be amplified without substantial selective amplification of some oligonucleotide sequences over others, e.g., due to competitive effects, unlike in many prior art techniques. Although some drift may occur during the amplification process, the drift may be relatively small. For example, in certain embodiments, the ratios or percentages of a representative unique oligonucleotide sequence, relative to the starting overall population, on the average, may change upon amplification by no more than about 10%, no more than about 8%, no more than about 6%, no more than about 5%, no more than about 4%, no more than about 3%, no more than about 2%, no more than about 1%, etc., relative to the starting ratio or percentage of the oligonucleotide sequence, prior to amplification. However, it should be noted that the oligonucleotide sequence itself, prior to any amplification, may also exhibit some variability, which is not included in the above numbers.

The unique oligonucleotides within a plurality of oligonucleotides may have the same or different lengths. If more than one unique oligonucleotide is present, then the unique oligonucleotides may independently have the same or different lengths. For example, in some cases, a plurality of oligonucleotides may have an average length (number average) of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides. In some cases, the average length may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides. Combinations of any of these are also possible, e.g., the average length may be between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

In one set of embodiments, any suitable technique may be used to amplify the plurality of oligonucleotides. In some cases, for each oligonucleotide to be amplified, at least about 100, at least about 300, at least about 500, at least about 1,000, at least about 3,000, at least about 5,000, at least about 10,000, at least about 30,000, at least about 50,000 at least about 100,000, at least about 300,000, at least about 500,000, at least about 1,000,000 copies, at least about 3,000,000 copies, at least about 5,000,000 copies, at least about 10,000,000 copies, at least about 30,000,000 copies, at least about 50,000,000 copies, or at least about 100,000,000 copies of the oligonucleotide may be produced using any of the amplification techniques discussed herein (e.g., including PCR amplification, *in vitro* transcription etc.). As discussed, in some cases, the amplification may occur without substantial selective amplification of some oligonucleotide sequences over others.

Any suitable technique may be used to generate the plurality of oligonucleotides. For example, the plurality of oligonucleotides may be synthetically produced, grown within a cell, grown on a substrate (e.g., in an array), or the like. Techniques for producing oligonucleotides are known to those of ordinary skill in the art. The plurality of oligonucleotides may also be computationally designed in some embodiments.

In one embodiment, the oligonucleotides may be amplified using PCR (polymerase chain reaction). In some cases, the oligonucleotides may be amplified while contained in a common liquid or solution. This is to be contrasted with certain PCR techniques, such as emulsion PCR or digital PCR, which requires separation of the oligonucleotides, e.g., into separate compartments or droplets, prior to amplification so as to prevent relatively selective amplification of certain oligonucleotides from occurring. However, surprisingly, it has been found that such separation is not required, and other techniques (such as is described herein) may be used to prevent or reduce selective amplification while keeping the oligonucleotides together within a common solution.

As mentioned, in some cases, by using certain oligonucleotide structures and/or certain types of PCR techniques, the amount of selective amplification that may occur may be reduced or eliminated. For instance, in one set of embodiments, oligonucleotides are chosen to minimize competitive effects. For instance, the oligonucleotides may have substantially the same lengths, and/or share identical or similar portions or regions.

For example, in one set of embodiments, the oligonucleotides may have a distribution of lengths such that no more than about 10%, no more than about 5%, no more than about 3%, or no more than about 1% of the oligonucleotides has a length that is less than about 80% or greater than about 120%, less than about 90% or greater than about 110%, or less than about 95% or greater than about 105% of the overall average length of the plurality of nucleotides.

In another set of embodiments the oligonucleotides may share one or more regions, such as index regions, that are identical or substantially similar. The oligonucleotides sharing index or other regions may have substantially the same lengths, as discussed above, or different lengths. In some embodiments, the oligonucleotides may comprise at least two index regions that each are identical or substantially similar, surrounding a variable region having different nucleotide sequences, and optionally, different lengths. For example, the oligonucleotides may include, in sequence, a first region that is identical or substantially similar to the other oligonucleotides, a second region that is not identical, and optionally, a third region that is identical or substantially similar to the other oligonucleotides. In some embodiments, competition of oligonucleotides may be controlled by using oligonucleotides selected to reduce amplification bias. For instance, in some cases, groups of oligonucleotides that have similar compositions may be amplified together.

In some cases, the index regions may have a length of greater than 5, 7, 10, 12, 14, 16, 18, or 20 nucleotides, and/or have a length of less than 30, 28, 25, 22, 20, 18, 16, 14, 12, or 10 nucleotides. For instance, the regions that are identical or substantially similar may have a length of between 18 and 22 nucleotides. The regions may be identical, or differ by no more than 1, 2, 3, 4, or 5 nucleotides (consecutively or non-consecutively) within the region.

In certain embodiments, primer sequences may be added to facilitate the PCR reaction. For example, the primer sequence may include sequences substantially complementary to a region within the oligonucleotides, e.g., an index region. A variety of such sequences suitable for PCR or *in vitro* transcription may be readily obtained commercially.

In some embodiments, the primer sequence may also include other sequences, e.g., promoter sequences or other sequences that may be added to the oligonucleotide

during PCR amplification, such as is shown in Fig. 1 with a T7 promoter. Accordingly, oligonucleotides comprising the original sequence and one or more promoter sequences may be produced in certain cases. Besides the T7 promoter, other suitable promoters that may be used include, but are not limited to, T3 promoters or SP6 promoters. Such promoters may be useful, for example, to facilitate transcription to produce RNA, as is discussed in more detail below. In addition, in some embodiments, more than one promoter may be added.

In one set of embodiments, more than one sequence containing a PCR primer may be used, e.g., to amplify different subsets of the plurality of oligonucleotides. If more than one primer-containing sequence is used, the PCR primers contained on each of them may be the same or different. Examples of suitable PCR primers include those described herein. Thus, for example, in one set of embodiments, for example, the plurality of oligonucleotides may include different subpools having different index regions or other regions as discussed above, which may be selectively amplified through the use of an appropriate sequence including a PCR primer. Any suitable number of subpools may be created for a plurality of oligonucleotides. For example, at least 1, 2, 4, 10, 20, 96, 100, or 192 subpools of the plurality oligonucleotides may be selective amplified, through the use of specific PCR primers.

Thus, in some cases, the oligonucleotides may be formed into "pools" or groups or sets of oligonucleotides within the plurality of oligonucleotides that share one or more common features, such as an index region or other identical sequence. For instance, the common feature in a group or set of oligonucleotides may have an identical sequence of nucleotides of at least 5, 7, 10, 12, 14, 16, 18, or 20 nucleotides, and/or less than 30, 28, 25, 22, 20, 18, 16, 14, 12, or 10 nucleotides. For instance, the common region that is identical or substantially similar in a group of oligonucleotides may have a length of between 18 and 22 nucleotides. The common regions may be identical, or differ by no more than 1, 2, 3, 4, or 5 nucleotides (consecutively or non-consecutively) within the region. In some embodiments, each group or pool may contain two (or more) unique index regions that are not used in any other pool, e.g., to reduce the contamination of off-target amplified products from the amplification products of another.

In certain embodiments, PCR amplification may be monitored, and controlled to reduce or minimize selective amplification. For example, in one set of embodiments,

real-time PCR techniques may be used. In some embodiments, the extent of the PCR reaction may be monitored or controlled, for example, by illuminating the solution containing the oligonucleotides with suitable light and determining the amount of fluorescence that is present to monitor the PCR reaction. Accordingly, for example, the reaction conditions may be controlled such that the oligonucleotides react in conditions that minimize the amount of selective amplification, for example, by providing an excess of nucleotides, ions (e.g., Mg^{2+}), enzyme, etc. Once the oligonucleotide concentrations have reached the point where competitive effects may start to occur, the reaction may be stopped before significant selective amplification begins.

After amplification as discussed above, the amplified oligonucleotides may then be transcribed to produce RNA. Further amplification may also occur in this step. For instance, in some cases, each oligonucleotide can be used to produce, on the average, at least about 50, at least about 100, at least about 300, at least about 500, at least about 1,000, at least about 3,000, at least about 5,000, at least about 10,000, at least about 30,000, at least about 50,000 at least about 100,000, at least about 300,000, at least about 500,000 or at least about 1,000,000 transcribed RNA molecules. In some cases, the mass of RNA that is produced may be at least about 10, at least about 20, at least about 30, at least about 50, at least about 100, at least about 200, at least about 300, or at least about 500 times the mass of the oligonucleotides. Thus, for example, one microgram of oligonucleotides may be converted into at least 10 micrograms, at least 30 micrograms, or at least 100 micrograms of RNA.

In one set of embodiments, transcription may occur *in vitro* by exposing the amplified oligonucleotides to a suitable RNA polymerase. A variety of RNA polymerases are available commercially, including T7, T3, or SP6 RNA polymerases. Other non-limiting examples of RNA polymerases include RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, or RNA polymerase V. The RNA polymerase may arise from any suitable source, e.g., bacteria, viruses, or eukaryotes. In some embodiments, more than one RNA polymerase may be used. In addition, as previously discussed, in some embodiments, the amplified oligonucleotides may include promoter sequences, such as one or more of T7, T3, or SP6 promoter sequences, that can be used to facilitate the transcription process. Those of ordinary skill

in the art will be aware of suitable conditions for causing transcription *in vitro* using RNA polymerases.

In some embodiments, the total amplification bias may be reduced by changing the relative amount of amplification produced by the PCR and the *in vitro* transcription. For example, the PCR can be used to produce smaller amounts of DNA than are typically produced in a PCR, to reduce the amplification bias of this process. However, this reduced yield can be compensated in some cases by increasing the duration of the *in vitro* transcription reaction.

The RNA may, in turn, be reverse transcribed to produce DNA. In one set of embodiments, reverse transcription may occur by exposing the RNA to a suitable reverse transcriptase enzyme. In some cases, the reverse transcriptase may be a viral reverse transcriptase, e.g., M-MLV reverse transcriptase, AMV reverse transcriptase, or the like. A variety of reverse transcriptase enzymes are commercially available. Those of ordinary skill in the art will be aware of suitable conditions for causing reverse transcription to occur.

In certain embodiments, reverse transcription may be facilitated through the use of primer-containing sequences, e.g., containing primers for reverse transcription. In some cases, the primer-containing sequences may contain other sequences or entities as well, although this is not necessarily a requirement. The primer-containing sequences may be added at any suitable point, e.g., just before starting transcription reaction. Suitable transcription primers for conducting reverse transcription may be obtained commercially.

In one set of embodiments, the primer-containing sequence may be incorporated into the DNA during production of the DNA by the reverse transcriptase. In some cases, the primer-containing sequence may contain other entities, and/or sequences suitable for attaching other entities (e.g., on the 5' or 3' ends, internally, etc.). For instance, the primer-containing sequence may contain a non-nucleic acid moiety, such as a digoxigenin moiety, a biotin moiety, etc. located on the 5' end, the 3' end, internally, or the like. In some cases, the signaling entity that can be subsequently detected or determined may be introduced to the DNA. For instance, the signaling entity may be fluorescent, or a specific nucleotide sequence that can be determined, e.g., enzymatically. Examples of signaling entities are discussed in more detail below.

In some cases, the RNA may be purified prior to reverse transcription. However, it should be noted that purification is not required, and in other embodiments, the RNA may be reverse transcribed to form DNA without any intermediate purification steps. If the RNA is purified, it may be purified using any suitable technique, e.g., by passing the RNA over a suitable column to remove oligonucleotides.

Optionally, in some embodiments, the RNA may be separated from the DNA or the DNA may be purified in some fashion. For example, the RNA may be selectively degraded, relative to the DNA. In one set of embodiments, the RNA may be degraded relative to the DNA by alkaline hydrolysis. For instance, the pH of the solution may be raised to at least about 8, at least about 9, at least about 10, etc. Any suitable alkaline may be used to raise the pH. In some cases, after degradation of the RNA, the pH may also be lowered, e.g., to about 7, to about 7.4, to physiological conditions, or the like. In some cases, techniques such as enzymatic degradation can be used to selectively degrade RNA, relative to DNA. The DNA may also be purified using techniques such as column purification, ethanol precipitation, and/or solid-phase reversible immobilization techniques. In addition, in some cases, the DNA may be concentrated, e.g., through evaporation techniques.

In addition, techniques such as those described above may be scaled-up or "numbered-up" to produce larger quantities of material. For example, a process may be repeated using multi-well techniques or by simultaneously running multiple reactions in parallel, etc. to produce larger quantities or masses of oligonucleotides. As a non-limiting example, in one embodiment, processes such as those discussed herein may be performed using multiple wells of a microtiter plate (e.g., having 96, 384, 1536, wells, etc.) to increase output.

The DNA may be used for a variety of purposes, in different embodiments of the invention. For example, in certain embodiments, the DNA may be hybridized to nucleic acid species in liquid samples, e.g., extracted from a variety of biological sources, including human. In some cases, the DNA may be used to physically separate one set of nucleic acids from another, or as primers for PCR or reverse transcription.

In addition, as previously discussed, in certain aspects, signaling entities are incorporated into DNA in some embodiments. The signaling entities may be determined for a variety of purposes. For example, the DNA that is produced may be used as a

biological probe, and the signaling entities may be determined in some fashion, e.g., quantitatively or qualitatively, to determine a characteristic or feature of the probe. Examples include, but are not limited to, the position of the probe, the activity of the probe, the concentration of the probe, or the like.

In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of the entities within the sample may be determined in two or even three dimensions.

In some embodiments, the spatial positions of the signaling entities may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No.

7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic

acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can

switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first

portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. . Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable

entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520,

ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl

Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13,

SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

In some aspects, the nucleotides can be used to study a sample, such as a biological sample. For instance, the nucleotides may be used to determine nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like. In some embodiments, for example, at least some of the plurality of oligonucleotides are complementary to a portion of a specific chromosome sequence, e.g., of a human chromosome.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some

cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

One non-limiting example of such as system may be found in U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.* incorporated herein by reference in its entirety.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.* In addition, U.S. Provisional Patent Application Serial No. 62/031,062,

filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.* is incorporated herein by reference in its entirety.

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

This example illustrates high-throughput hybridization construction of DNA probes, according to certain embodiments of the invention.

Overview. This protocol uses complex libraries of oligonucleotides as templates for the enzymatic construction of large quantities of single-stranded DNA molecules that can be chemically labeled and which are designed to be hybridized to specific sets of nucleic acid targets that can vary significantly in complexity, i.e. the number of unique target sequences. The use of this protocol involves several basic steps: 1) computational design and optimization of a set of oligonucleotide sequences that will serve as the hybridization regions on the sample of interest; 2) computational design and optimization of a large set of short oligonucleotide sequences to serve as highly specific PCR primers; 3) computational construction of the template molecules; 4) synthesis of the template molecules to create the template library; 5) selection of a sub-set of *in vitro* template molecules from a template library via PCR; 6) *in-vitro*-transcription-based amplification of the *in vitro* template molecules into RNA, which serves as the final template for the hybridization probes; 7) reverse transcription of the RNA back into DNA using chemically-modified primers; 8) removal of the RNA via alkaline hydrolysis; and 9) purification of the chemically-modified ssDNA molecules, called probes in this example. These steps are discussed in detail below.

Construction of Probes. Selection of Hybridization Regions. The software OligoArray 2.0 was used to design a large set of potential hybridization regions in the *E. coli* transcriptome in this example. Briefly, this software selects hybridization regions that lie within user-specified ranges for length and melting temperature. This software also screens for off-target hybridization as well as potential secondary structure. This software was used to generate hybridization regions of 30-nt length, a melting-temperature range of 80-85 °C, a GC content of between 50-60%, and a secondary structure melting threshold and cross-hybridization-melting temperature of 75 °C using all annotated, transcribed mRNAs in *E. coli* (K-12 mg1655; NC90013.2).

Design of Hundreds of Orthogonal Primers. The general hybridization probe set may require far fewer unique sequences than is provided in the complex sets of oligonucleotides typically generated by array-based synthesis—the *oligopool*. To exploit this disparity in complexity and significantly lower the cost per experiment, this example uses a method to embed a large number of unique template sets within a single oligopool. Briefly, each template molecule was flanked by a unique pair of PCR primers common only to the template molecules for all probes within a given set. To facilitate the embedding of hundreds of unique probe template sets within a single oligopool, a protocol was created for constructing hundreds of orthogonal PCR primers.

Specifically, the protocol starts by truncating the members of an existing library of 240,000, semi-orthogonal, 25-mer oligonucleotides to 20-nt length and selecting oligos on parameters optimal for PCR: a narrow range of predicted melting-temperature (65-70 °C) and GC content (50-60%); the absence of contiguous runs of the same base longer than 4, i.e. AAAA; and the presence of a 3' GC-clamp, i.e. 2-3 G/C within the final 5 nt. BLAST was then run with these optimized primers against all annotated RNAs in the *E. coli* transcriptome as well as the T7 promoter (TAATACGACTCACTATAGGG) (SEQ ID NO. 1) and a common priming region (P9: CAGGCATCCGAGAGGTCTGG) (SEQ ID NO. 2) and potential primers with hits to the transcriptome with 12-nt or longer of homology, or with a hit of any length within 3-nt of the 3' end of the potential primer, were removed. Finally, the remaining primers were screened for homology to each other, again using BLAST. Any primer with longer than 11-nt of homology or with any homology within 3-nt of the 3' end of another primer was removed. These cuts reduced the original 240,000 oligos to 198 highly optimized primers for the *E. coli* transcriptome. The final set of 198 primers are listed in Fig. 3.

If required, more primers could be generated using established techniques to create a larger set of initial oligonucleotides, or by relaxing the stringency of the cuts described above. Finally, by changing the transcriptome used to screen the primers, this approach can also be generalized to the generation of optimal index primers for any organism.

Template Construction. To design the template libraries used to create our probe sets, the desired RNA targets were first selected to be stained simultaneously. Individual

template molecules were designed by concatenating the following sequences: i) the first of two unique primers for the appropriate mRNA group, ii) the common primer P9, iii) the site for the nicking enzyme Nb.BsmI, iv) the reverse complement of the hybridization region to the target, v) the reverse complement of the nicking enzyme Nb.BsrDI, and vi) the reverse complement of the second unique primer for the given mRNA group. Fig. 2 demonstrates this organization. Multiple probe template sets were combined into large oligopools and these pools were synthesized via CustomArray.

Fig. 2 shows an example template sequence containing a probe the mRNA, acnB. Underlined at the beginning is the sequence of the first primer, not underlined is the common P9 priming site, then underlined is the Nb.BsmI site, not underlined is the reverse complement to the hybridization region for acnB, next underlined is the reverse complement of the Nb.BsrDI site, and the final not underlined portion is the second unique primer. This template is one of 736 used to create probes to stain all mRNAs expressed in the copy number range of 1-10 per cell transcribed from the *E. coli* genomic locus corresponding to base pairs 1-100 kb. See Fig. 4 for the sequences of the 736 probes.

Index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. 0.5 to 1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired sub set while the reverse primer was a 5'concatenation of this sequence with the T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed before the plateau in amplification, often at concentrations about 10-fold lower than would correspond to this plateau, to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the

manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 micrograms of template DNA was amplified into 100 to 200 micrograms of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 2 to 4 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1-2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNPT, 1-2 nmol of fluorescently labeled P9 primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of a solution of 0.25 M EDTA and 0.5 N NaOH was added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 100- microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspend in 10 microliter RNase-free water and stored at -20 °C. Denaturing poly-acrylimid gel electrophoresis and absorption spectroscopy revealed that this protocol typically produces 90-100% incorporate of the fluorescent primer into full length probe and 60-75% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol

of fluorescent probe. The small reaction volumes were conducive to the use of high-throughput fluid handling techniques and significantly lower the cost of this reaction as compared to alternative approaches. Thus, 24-96 probes could be constructed in parallel, with minimal hands-on time, across two days, for a final cost of ~$14 per 2 nmol of each probe set.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A

or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

ABSTRACT

The present invention generally relates to systems and methods for producing nucleic acids. In some aspects, relatively large quantities of oligonucleotides can be produced, and in some cases, the oligonucleotides may have a variety of different sequences and/or lengths. For instance, a relatively small quantity of oligonucleotides may be amplified to produce a large amount of nucleotides. In one set of embodiments, oligonucleotides may be amplified using PCR, then transcribed to produce RNA. The RNA may then be reverse transcribed to produce DNA, and optionally, the RNA may be selectively degraded or removed, relative to the DNA. In one set of embodiments, the oligonucleotides may be chemically modified. These modifications may include, but are not limited, to the adding of fluorescent dyes or other signaling entities.

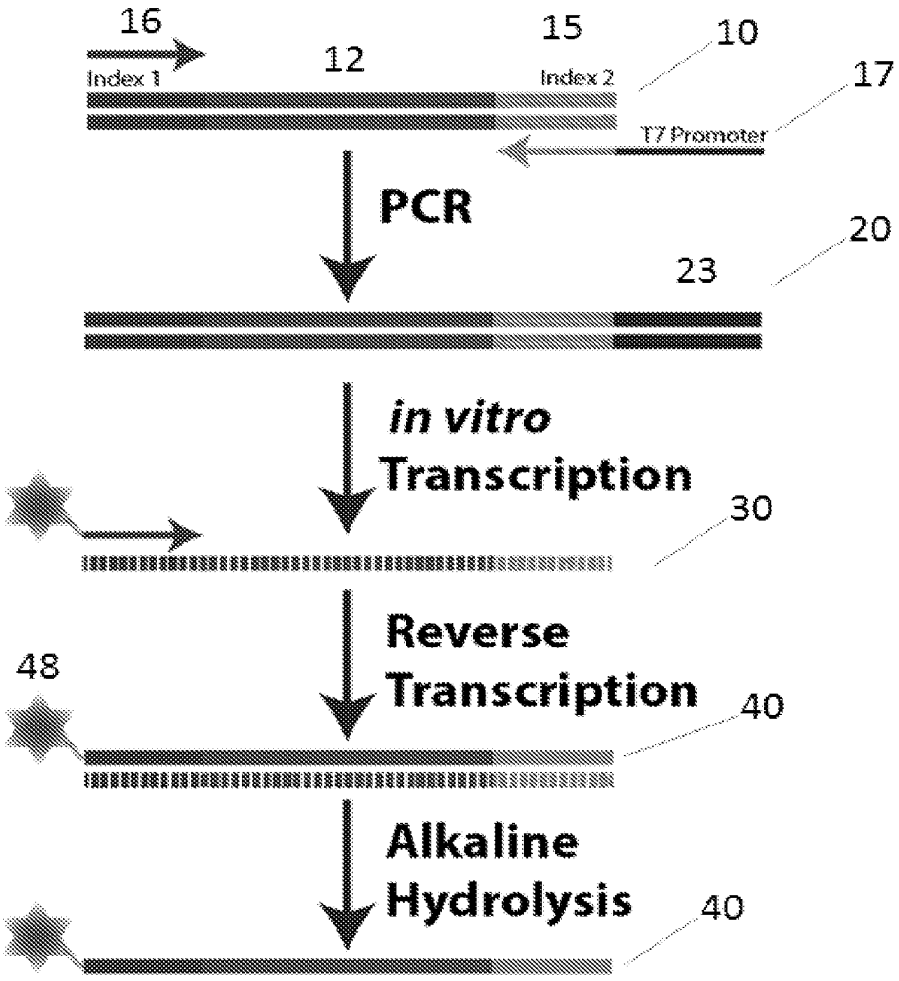


FIG. 1

ACAACCGCGTGTTACAAGGCCAGGCATCCGAGAGGTCTGGGAATGCC
AGCAGCTCTACAAGTGCGGCCATTTGGTTCATTGCGCTTCTCCGGAAAC
ATTAGC (SEQ ID NO: 3)

FIG. 2

FIG. 3

>bc25mer_8982-20
CGGTAGGACCCTACACATCC

>bc25mer_37497-20
TGTGGCGCCTAACCATACAC

>bc25mer_48272-20
GACACCGAGCTAGAATTCGG

>bc25mer_96434-20
TGTCTTGGGTCCTCTTACGG

>bc25mer_135960-20
TATCACGTGAGGCTCCGTGT

>bc25mer_217459-20
GTCTACGGATCCTTGCGTGT

>bc25mer_5177-20
ACGGGTACATTGGATCCACT

>bc25mer_7835-20
AAGACGTGGTCCACCCTGAT

>bc25mer_13747-20
TGCCAGGTCCACTACTTGGG

>bc25mer_14156-20
GACCCGAAGGAGAATGCAAA

>bc25mer_14649-20
ATGTCGTCTCTCAGGGCTTT

>bc25mer_16849-20
AAGGGCGGAAGCACTCTCTT

>bc25mer_25227-20
TGGTCCCGCATTTGTACCTT

>bc25mer_45585-20
AAGTCCGTCTGCTAAGCCAT

>bc25mer_46295-20
CTTCCAGAGTCCTGGCATGA

>bc25mer_46396-20
TCTAACGAGCTCCGTGGTTG

>bc25mer_55377-20
CCTGGGCAGGAAACAACACT

>bc25mer_57092-20
GCGCACACCTTAGTCATAGC

>bc25mer_69827-20
GTGCGATTGTCACATGTGCT

>bc25mer_85616-20
GGCTTGTCCTCAATTATCCG

>bc25mer_90788-20
CCCACGTGACTTGACTTCTC

>bc25mer_108211-20
CCTCTACATGCTCCGGATCC

>bc25mer_113656-20
GAATAGTAACCCGGTCGCAA

>bc25mer_121326-20
TGGGCCACTAGTCTTCGCTA

>bc25mer_123881-20
CGGCACGTACCCTTTACAGG

>bc25mer_138428-20
GATCAGTGAGCCTCACCAAG

>bc25mer_140203-20
TGGTATGTGAGCTCTTTGGG

>bc25mer_159878-20
AGAGTCACGCATACTTGGGA

>bc25mer_163235-20
GGGTCAGGCGTCATAGATTG

>bc25mer_184954-20
AGAATGTGGAGGACATGTGG

>bc25mer_188166-20
TCGGACTGTGAGACATTTCC

>bc25mer_192833-20
ATCAAATGCCTCGGTCGTTG

>bc25mer_193757-20
GCAATGCTATGCTGGGACAA

>bc25mer_199694-20
ATGGAGGCCGTTCTCTGACA

>bc25mer_200273-20
GCAGATGGGTGCATTCATTC

>bc25mer_208214-20
TATGCACTCTGCGGATGACC

>bc25mer_218789-20
GACCCGGGAGTTATGATACG

>bc25mer_235702-20
CTCGGGTAGGCTATCTCCAA

>bc25mer_239358-20
CTGCAGGTGCTCTAGATGGA

>bc25mer_5878-20
GTCCAGGCTCGTCTTCTCGA

>bc25mer_9018-20
CTTACCTGAGCGCAGTTCGA

>bc25mer_9904-20
GTTGTTCGATCCCTCCACCA

>bc25mer_10475-20
CGGCAGAGGATAATCCTAGC

>bc25mer_11594-20
GGGAGTCCGGATGTTAGTGC

>bc25mer_12189-20
CGCATCTATGCTCCGCTTAC

>bc25mer_14363-20
AACGCGCTACCCAATTCTAG

>bc25mer_16178-20
TACTGGACGATTCCCGACTG

>bc25mer_17067-20
AGCGTCGTACCCAGTTAAGC

>bc25mer_18548-20
GAGATTCACGCCCTCATGAG

>bc25mer_20080-20
CCGCTAGCGTTACCTCTACC

>bc25mer_20601-20
CCACACCGTATGCATTTCTG

>bc25mer_26252-20
CCTTAGCGCACAAAGAGACG

>bc25mer_28353-20
GTCCTGATGGCTTTCTCACG

>bc25mer_30927-20
GGTCGTCGGTAGATCATTGC

>bc25mer_32240-20
AGAGCCGGTATGATCCATCG

>bc25mer_32875-20
CCGGACGAAGGTTGATATCC

>bc25mer_39270-20
ACCCGGTACCTGTTATCACG

>bc25mer_40398-20
ATCCCGTCTCGTTCTATGGG

>bc25mer_41429-20
CAGGTGAGGGTCTCTCCTTC

>bc25mer_42283-20
CCTTATGGAAACGTGATGCC

>bc25mer_43506-20
GGCTGGGCATACATAGACCT

>bc25mer_45853-20
AGTGGCGGCTATTACCGGAT

>bc25mer_47706-20
TGATAACGTCCGCTCGTTGA

>bc25mer_54028-20
TACATCGTTGAGGCCCGTTT

>bc25mer_54772-20
CGTTGTGGCATCAGCTAGAG

>bc25mer_55112-20
ACTGGCCCACACACTTACCT

>bc25mer_57695-20
ATCGGGTCACGGAATATGAC

>bc25mer_58256-20
CGTAGGTTCACAGATTTGCG

>bc25mer_59008-20
TAGGGCACGAGAAGGTATCC

>bc25mer_61977-20
ACTAGCTTGTATCGCCGGAT

>bc25mer_63426-20
CGGCGGTTTGATATTCGAAG

>bc25mer_63840-20
GCCGAGTGTTTATGAGCAAG

>bc25mer_65782-20
TCTCCGAAGATCCGATATCG

>bc25mer_66426-20
GATGCAGGGCCTAATTAACG

>bc25mer_68963-20
TGACGCGGCTAAATACTGAC

>bc25mer_70828-20
GCAATGCACGCTCTCCTAGT

>bc25mer_71011-20
TGGCGTCAGATGGATTAGGA

>bc25mer_71626-20
GTTGGCGGGAGACTAAGAAG

>bc25mer_72701-20
TCGCGTCTCGTCCTTCTACC

>bc25mer_79345-20
CGGTCACGTCGGAAATAACC

>bc25mer_79641-20
ATTAACGGGCCAGGTTACTG

>bc25mer_81698-20
ATTGTATGGAGGCGCCCTAT

>bc25mer_82313-20
GCCAGCTTTCCAAGATTCAG

>bc25mer_83004-20
ACAACCGCGTGTTACAAGGC

>bc25mer_83517-20
GCTAATGTTTCCGGAGAAGC

>bc25mer_83687-20
GAAGTATCCGGCATCACAGC

>bc25mer_85062-20
TCGTTACGGACTTTCACGAC

>bc25mer_87436-20
GCCATTCTGACATACCCAGA

>bc25mer_90109-20
GTATTTGAACCGGCCAGCTG

>bc25mer_95782-20
GGTTGTGGGCCATATCCAAT

>bc25mer_96522-20
GGCGTTCCATCGAAACTCTA

>bc25mer_96896-20
GGCCATCGACTTAGATTCCA

>bc25mer_97519-20
GGTGCTCCTGCCATTATAGG

>bc25mer_100107-20
TGTCGACGTGCGGAAAGTAG

>bc25mer_100109-20
GCCGCACGAGTATGCTACTG

>bc25mer_100502-20
TGCTGCCGCTTACTACTGCT

>bc25mer_102626-20
TAGCACTCTTCTGGCCATCG

>bc25mer_103544-20
CGGTAATCGGTTCACCAGTG

>bc25mer_108258-20
TTTAGTGCCTGGCTCCGTTG

>bc25mer_109838-20
GGAGTTACAGCCACTTTCGG

>bc25mer_114408-20
GTGTGGACGGAACCTGACAA

>bc25mer_115589-20
CACACGTGGAATTGTTCTGC

>bc25mer_116534-20
CTTGTACATGGAGGGCGACT

>bc25mer_116605-20
AATCCGGTGTACAGGTTCCC

>bc25mer_121894-20
AGGTGCGACGATATACACCA

>bc25mer_123748-20
ACTGCATATGACCGCTGCAA

>bc25mer_125347-20
CCTTCCATTAGCCGAGATCA

>bc25mer_125541-20
GACGCGAGGAGTGATCGACT

>bc25mer_126017-20
CGACCCTCACACATTTGGGT

>bc25mer_126439-20
CGCTCCCGTATTCTTCTGTG

>bc25mer_128324-20
CAGGGATAGTCAGGCCAGTG

>bc25mer_130576-20
CGTAATGAGTGCTTCGCCAT

>bc25mer_132438-20
CTCGTCTCGTGGAACACATG

>bc25mer_133420-20
TATGCTGGGCAGTAATCAGG

>bc25mer_133575-20
GCATCTCGTAGCATCCTGCT

>bc25mer_134218-20
AGTATGGTACCGGGAACAGC

>bc25mer_136989-20
ACCATTCTCGCAACTCGCTA

>bc25mer_137798-20
CGTGGGCGAAGTACTTGGTC

>bc25mer_138338-20
CCGGACGTCTTCGATAATGC

>bc25mer_139329-20
TGATGCTCTTTGCAGTTCGG

>bc25mer_139807-20
ATGCCGGTCATAACAGTGTC

>bc25mer_140760-20
TGGTGGCTCGTTATCACAAG

>bc25mer_141708-20
TCTGAAGCGTGGCCATTACC

>bc25mer_141799-20
CATATGCCGGACATTCAGCT

>bc25mer_142929-20
ACTAATGGTCCTGCGGCATA

>bc25mer_146048-20
GGTGCCGTGTTGCATGTAAG

>bc25mer_149885-20
AAGCGGTCGTGGTTTATACC

>bc25mer_150249-20
GCCACCTTGTATGGTATCGA

>bc25mer_150996-20
TAATGCTTAGGCCCGTCGGT

>bc25mer_153952-20
GACGCGACGGATTATTTAGG

>bc25mer_156264-20
GGTCGCGCCCATATATAAGG

>bc25mer_156364-20
TGACTACGGTTGGGTGCATC

>bc25mer_157007-20
TCAGGCGCTCATTGTATGTC

>bc25mer_160860-20
GTAGATACTCCCGGCCCGAA

>bc25mer_161471-20
GCAGCACTTAGGGCAGCATC

>bc25mer_163904-20
CGCGGGAACCATATTAGGAA

>bc25mer_164195-20
GATCCATGGCCAGTTCGTAT

>bc25mer_164787-20
GTGCGCCAAAGGACTTAGTG

>bc25mer_166664-20
GATCCCGCTATTCACCGATT

>bc25mer_167568-20
GCATAGATGGTTCACCCGTA

>bc25mer_169085-20
GTCTATAAGCCGCGCTGCAA

>bc25mer_170764-20
GCGGAGTATGCCATCATGAG

>bc25mer_171455-20
TGCCGCGATCATCTACTATG

>bc25mer_174112-20
GACCGGGTATTCGACGTCAT

>bc25mer_175501-20
GCAGTGCGGGTAGATACGCT

>bc25mer_180358-20
AGGCGTGGGTAGCAACGTAT

>bc25mer_180797-20
GCTCATTGGACTTTCTCCCA

>bc25mer_181432-20
AAGGACCACGTATCTGCATG

>bc25mer_181702-20
ATCTGATAGTGCCGCGACGT

>bc25mer_182418-20
TGACCTGCACGGATAGTAGG

>bc25mer_182774-20
GTTATTCGCAGTCCTTGGGT

>bc25mer_183321-20
GTGCAGTGTGCTTAAATCCG

>bc25mer_188636-20
CGCGGGTGTTAAATAAGGAC

>bc25mer_189039-20
GCGAGGCGTGGTAATAGTCA

>bc25mer_190422-20
CAAGGCGCAAACATAGACAG

>bc25mer_191525-20
GGAACAAGGGCGTCTATGTC

>bc25mer_193359-20
GCAGTGCGGATAAGCTACAC

>bc25mer_193734-20
AGAGATGCGTGTAGGCGATT

>bc25mer_194512-20
CCCTTTCGAGCTAAGTTTGG

>bc25mer_196660-20
CCTTAAGCAACCCGTCGATG

>bc25mer_197001-20
TTGCTCGTGACTGAACAACG

>bc25mer_197794-20
GGTCGTGCGTATAAGCCTCA

>bc25mer_198508-20
CGGATGGTCTTCGTTTAACC

>bc25mer_198891-20
GGCACGGTGGCTAGTAACGA

>bc25mer_200581-20
GCCGGCCCAACTGATAGTAG

>bc25mer_200721-20
TCGACTTCGGTCACCTTTCC

>bc25mer_206382-20
CGGGTCGATACTTTCCTCGT

>bc25mer_207645-20
GTGCCGTGTGTAATATCCGA

>bc25mer_210956-20
CGCGCCTGGAATAACTCCTA

>bc25mer_210958-20
GTCAGGGAACCGTTTCTTCA

>bc25mer_210961-20
CTTGTCATGTACCCGAATGG

>bc25mer_211638-20
CGACCTGGCTACGTAGAACC

>bc25mer_211770-20
CTAGCTAAAGGGCCGTGCGT

>bc25mer_213351-20
AAGTATTAGCGCGGCAACGT

>bc25mer_213506-20
GGTTGCGTGCCACTTAAAGC

>bc25mer_216161-20
GGAGGTTCGGTTGTACTGCA

>bc25mer_216486-20
GTTGTCGTCCTCCATCGGTT

>bc25mer_218887-20
GGGACGGAACTACACATGTG

>bc25mer_221408-20
GATCCATCCTGATTGAGGGT

>bc25mer_222754-20
CTTCATGGTACCGGTTGAGA

>bc25mer_223329-20
TCGGTCGGCTGTAAGGATAC

>bc25mer_224422-20
ATTGCACGAGGTCAGAGTCG

>bc25mer_225143-20
GCTTAGATCCGCTCGCTACG

>bc25mer_227128-20
CGGCTGGCCTACTGTAGAGA

>bc25mer_227360-20
GCAACATGACCTGTCATCGC

>bc25mer_228435-20
GGCCGCACGATATATTTGAC

>bc25mer_229488-20
CAGCCGGGCCATAATAGTTG

>bc25mer_229709-20
TGCGGTGGACCTATTATCCT

>bc25mer_229929-20
CCTGCGTTAGGCAATCCATC

>bc25mer_230898-20
GTGGGCTCTTCGAAGTAACC

>bc25mer_233755-20
GCCGTGGACCACTAAAGTTC

>bc25mer_234401-20
GCTTAAGTCATGGGCGCATC

>bc25mer_236515-20
ACAGGTTAGTTCCGCGCACT

>bc25mer_237582-20
GGCGTGGCATTTAGACTACC

>bc25mer_238816-20
ACTACGGCCAACAACCAACA

>bc25mer_238839-20
GTTGACAAGGCTCTGTACGG

>bc25mer_239304-20
TTGTTCTCTTGCCGGTCGAT

>bc25mer_239849-20
GTTGATCGTAGCCAATTCGG

FIG. 4

Exp-1_acnB-73_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCAGCTCTACAAGTGCGGC CATTTGGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-191_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTCGCGATAGCAGCCAGGA AGCCTGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-671_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCGGGAAACCTTTCTGTT GCAGAGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2404_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCAGTTCCGCAGAAGCCAG GAAGACATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1307_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTGTGGCAGAAAGACTGCA TCACCAGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-779_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCGCGTTTGTTCGGCACAT GTGGAATAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1343_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGCGTGTTCACGTCAACTG GCTTCGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1961_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCTGCATCGGCTTCCAGCA GCTCAGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2131_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGCACGGAAGTGACCGAT GTTGGTCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2039_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCGGGTCGTTCGGAGCAC ACAGGATTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2262_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACGCGCACCACTCTTACCG AAGACGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2542_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGGCTTTCTCGGTGTACTG AGAAAGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2369_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTACCCAGACGGTTCGGGA AGTTACGGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1811_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTGTTCAGCTTGATGGTAC AACCAGCGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-428_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCGCCCAGGACTGCATAA CCTGCTTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-744_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCATAAACCACAGAACGGAG TTAGTGGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-393_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGTTGCCTGCTTTCGCTTTC TCTTCTACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-530_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTGCCGGAGAAAGGTCAT CGGTGTTAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1378_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGCCACGGTTCATAATGAA GTCCGGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2170_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCGGCAACTGACCTTTATG CGCATCCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-943_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTTCGCCGGTTTCGTGGTT ACGCACTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1881_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCCGTAACCTTCCGCGATC ATCCACTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-634_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGATCGGACCAACAACACC AGGCTGGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1224_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTGGTGTCCTGGGAACCT ACAGAAGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1768_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGACGCATCGGTTAGCTCAAA GGCCTGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2473_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTACCTGCGCCACGTAGGT CTGGTACTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1005_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACGGCCACCAGCACGCACT TCATCAATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-352_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGAACATCAGCAGCGTGTG AGAAAGTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2002_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATATCCGCCAGATCGATGTC GATCACTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2088_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACACTTCGTCGATCTTCTCA CCCTGTACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1681_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGCCTTTCTTCTCAACGGT CAGCAGACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1916_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATGCCCTGAATACGACGTT CCAGGGTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1478_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAAACGGGTATGGGAGTCAC CACCGGTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-706_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAGAACCCGTACCCACAAC GTCACCGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-897_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTCAATCACGTCGCCCATG TTCAGGTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1515_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAGACCAGAACCCGCCGGG AAAGAGATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1581_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGAAGCGCACCAGAACGGAT TCCGGCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2226_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTTCTTCGGTCAACTGTGCG GCGTCCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-2297_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTACCCATACACAGGGAAC AGCCAGGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1733_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAGATCCGGCAGACCTTCAA TTTCCAGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-305_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTTTGGCATCATCCAGCG CGTCGATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1266_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGGCACGCCAGGTCTTTC AGTTCATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1040_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTGGTGGTCAGGCCACGCC CGATAATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1646_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGATCGCATACAGCGGAATAG CGTGTACCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-848_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTGCACCCGCGTCTTCCA TCGTGTTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1436_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCAGCATACGGTTCAGCC AGGAGTGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_acnB-1105_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCTCAGCGACATCTTTCGC CTGACGGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-646_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTTCAAGGTCGCCCGCCAG TTTCTGTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1951_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGGTGCTGGAACCACCGGC ATAGAAGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1354_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCCCGTTGATACCAACAGC ATAACCTGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-686_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAAACGGGCATAACCGCGAT CCAGATAGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1189_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGAACACGTTGGGTATCGGT ATCGACAGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1410_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGTGAAGTACGGGTTGGTT ACCGACAGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-857_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGCTCACCCGGCTCGATCT TAGTCAGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTTGGTATAGTCGGACAGGT CGGCGTCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-85_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCGACACGCTGAAGGCCTTC GAAATGAAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-564_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGAAATGAGAGATCAGTTCG TCGGTGGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-508_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCTGGATTTCAGCTGACAC ACCTTCCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-967_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGGCATCGTTAATTTCGGG CATCGACTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1989_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTGCTTTCGGACCAATGGTA TTGGACTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1224_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCTTGTAGACGACATCAACC TGGTCCGGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1787_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCCGATCCAGGAATGGTCA CTTTACCGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-50_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACTACGAACCCTTCAGCAC CGTATACGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1101_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCCATGCACCTTCCATCTGA CGCATTTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1560_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGACCTGCACGCAGCGAG TTATATTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2382_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACTGGAACTGTTCTGCCTTG TCTCCATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-145_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTTCATCATTAACCGTGTC GCCTGTGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1700_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCAACCATAGTTGAACGTGA AGTCGTCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2299_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATTGTAATGCGATACCCGC AGACATACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-611_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATTTACGATCGCCTACCACGT TCCACCACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2229_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCTGGAATCCCAGTTTGTA TCCCAAACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-423_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCTGGCGCTATATTTACCG ACGCTGTAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2095_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGCCTACAGCATCATCCGA TTTACACAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-721_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTGGCGTCAGACTGACCTG GGTAGAGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2059_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGCCGTCCTGAGTCGCACA TTCGTAATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1857_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAACCCATTTGTGATCGTCA TCGATCGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1630_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCCATAGAGTACAGATAACG CCACATCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1260_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACCAAAGTTGAAGCTACCG GTGTTGCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1445_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATAGAAGAGACGACCACCGA GGCTTACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-892_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCATCTTCCATCTTGGTCAC TTTGGTGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1029_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGATCTTACGCACGTAGAAA CGGTTACCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1066_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGACGGCATCTTTCGAGGT ATCGTTACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1154_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAAGCCCAGACGATTCAGAC GCTCCTTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1751_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGACACGTGAACCATCTGTCG GGAAGTAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-15_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCTAAACAGCAGCGACGCT ATGAGCAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1595_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGGCTGCATGTTGGACAGGG AGTTATGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1525_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGGAAGCCCAACGTCACGTC TGTACCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-2024_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGATCATAATTACTGGCCT GATGCGGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1665_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGCTGTTATCCTGATCAGAG GTGCTCGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-234_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCTGAACCAGAAGGGTATCA CCATCACGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1911_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCATCTCTTTGCCGCCTAAA CCATCACCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-1318_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTTATCCTGCTGCACACC AGCCTGGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-932_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGATAGGCATAACCATAGC GACCGAGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-803_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAGGTTGCCGCTCACTTCAA CGCCAGAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_bamA-199_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGAACATCCTCAAAGTTGCC GGTAGCAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-721_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTGGGCTTGCACGTTCAAA CCAGTCACT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-764_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGTGAAAGCGCATTACAGA GCTGTTGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-418_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCGCTTTCAATTGCGCGTA CTGATCCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-143_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCTTGTGGAGGATAGTCGG AATAGCTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-637_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGAATTTGGTCCGGTCCGCC TGTAATGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-595_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAGCGTGCTAACACCTGTTC GCGGCTAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-240_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATTACCTCCACGCCAGGCA ATCACCAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_btuF-87_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAAAGGCAAGTTCAGTGTTG GCGGGAGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-349_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATATCGCGGATATGCAGCAG CCAGTTGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-202_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTTCAGGTCAGTGTGAAT GACCAGGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-167_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATTACGGTGAACAAAGAGTT CGCCCGGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-630_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGTGGTTGGCGTGTTTCAGC TTGAGGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-132_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTTAAACGTTCTGCAGGA ACGCGACTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-434_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGTTCCATGACGTTCAGTT CACACAAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-237_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGAGTACATCCACTGCATAC TGAACCACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-480_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTTCCACGCTGATTGCATA ATGGTGGAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-595_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCGTGACGGTAACGTTGCTC AAGGGTTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-279_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCCGCAACCGTAGTGGCCA CAGATAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_can-559_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGGCGGTAACATCCAGATC ACGCAGCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-414_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATGATTCTCGTCATAGTGC CAGGCCCGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-37_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGAAACAACGCCGCGATGAC GACGGGTAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-87_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAGCATGCAGACCACCAGC GTTACAATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-319_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGTTACGCCAGATTGCTGC GGAACCTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAGCCCAAAGTGAGATTTC AACCAGCGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-809_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTAACAACAAGCAAGCAAAGA CCGGTACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-449_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGATCATGACATAGAGCAGCC ATATTGCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-749_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCACCGTGTCCTGGAATTA AATGACCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTGCCAAACATATATGCGCC GGAGTCAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-575_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCTGCAGTAGCGAGTCCAC CGATAAAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-700_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCTTAAACATACTCTCGGT CAGATCGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-179_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATCAGCGCCAACAATAACC CGCATAACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-618_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGACGTCGAGATTCGCCCAC ATGCCATAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-282_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCACCAGCAATAGCGCGACA ATCCACCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_cdsA-665_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGAGGCTAACGCTGCGACAA TAGAGCAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-653_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAACGCGACCGTAGTGGATTT CGCCGGTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-93_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGAATCCAGCAGGGCGATC ACCTGATTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-384_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCCTTCATCAACATATGCG CCGATGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-148_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAACCACTGATGCGTCACCCA CTGACCGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-51_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTAACGGTGTCTGCATTGGCT GGCGTGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-724_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACCGCACAGTAGAGGCTGTA TTTGCCATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-794_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTCGATGGTACGCAGCAGTT CGTTAATGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-613_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATACGGGTGCTCTGACCAAT GTATACGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-540_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTTCAGAGCGCGCGCCGATG AAGCAATTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-348_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGGCATCAGCACGGTGTTA CGGGCAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-257_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGTCGTAGTCGGCGAATT TCATCGGCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-292_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCACAACGCGGAAGCCTTC TTTCTGGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_dapD-419_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACAAGAACCGACGGTCGCCC AGGTATCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-302_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGTAAACCGGCACCTTGCC ATAGAAGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-95_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCGACAGGATACGCGTCAA TACCGCCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-814_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGTGGCTGGTCATACCCGG TGAGGTATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-766_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTCCATCGCTGTCCAGCAT AACGTCAAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-614_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCCTCATAATCATAGAAGG TTCCGGACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-726_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTTTGCAACCTAACGTCGTC CATGCTTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-168_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCGACCGTGTAGCGCGATA AACACTTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-435_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACATTCCCACACTGGAACCT TCGCGGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-559_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCTTCACCGAGTATCGCAAC CGTGAACTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-394_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACCGGCAAACCCAGAGCAGA AATTTCTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-691_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATGCCTGCAAATTGGCCTC TTGTGACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-259_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCCATTGAAAGCGCAGATGC CATCACTCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-891_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGTCCGCCAGTTCCAGAATT CGTACTACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-133_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCCATCGACTTCAGTTGCGT CACGTCGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-55_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCTAACACCGCTGCGCCAGA ATTCAGAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ddlB-224_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTATAAGGCAAGCCCATCA GCTCGAGCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1219_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGTTGTTCACTACCACGCC CTGATCTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-703_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTCAGGTTAACCAGCGCACC ACCGGAGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1114_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCAGTTCCAGGTTCACGTT AACCTGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-948_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCGGAGGAATTAGGCAGA ACCTGGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-13_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTCAGAGCCAGTGCACTCAG TGCTAATGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1184_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTTGCTCATCTCAGCGCCTT CAATGCCGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1359_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAATGTTGAGTGCCAGCACA GACGGTTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-319_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGGAACCCAGCGCCATGAA TTTCTGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1149_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGGAGCTGGAATCAACCTGA TTCTGGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-852_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATAATACCCAGCTCACCGCGT TTCACCTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1015_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAAAGCTGCTGATCGGCTT ACCGTTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-904_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGGGCGTCAACTTTCATCGC TTTCGCCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-86_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATCTGCTGGGCTGTCGTTG CTGAAGAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-603_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCGCAGAGACAATCCCGGAA GTTACCGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGGTTCTGGATTTGGATCAG CGCGATATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-741_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGCGAGGATCGCGGTGTTG ATACCGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1394_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTGCATTAACAGGTAGATGG TGCTGTCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-371_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAACAACGTGGTTGTTGGTGA CGACATAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-411_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCCATCGCTCAGTTGAACT TTAATGACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-141_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTTAATGCTGACCACTGAA GGCATCACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1294_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACTGCCTGCTGGTTCGCGCC AATAATCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-528_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCGCAGTGCATCAGAATCC GCCATCTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-565_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAAACGGGTTACCAATCGC TACGGTGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-177_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCATACGCGGCGTATTAACG GTTGTGCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-1076_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTAAGCCCAGGGTCAGTT TGCTGCCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGATCTTTGCCAACCATCT TCGCGTCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-258_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCAGAACGGAGAGCTCTGG AACGGAGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-223_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGGCAGAACGGAGAATCATC ACCGAAGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_degP-48_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGTTGCAGAGAGCGGAGAT AACGCCAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-158_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGATGGTCATATCGCCTTCGT TCACCTGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-63_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGATTCGGGTTATCTTCGTCA GCGATCAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-314_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGATACTAAAGGAAGAACCGC AACCGCAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-28_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACTTTGTTGGCTGCTGCGTC GGTAAACTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-238_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCTTCGGTATAATCAACGGA ACCGCCGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-277_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTCGGGTTGGTCACGAT GAAACGAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_erpA-123_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGGTGAAACCATACTGGAAG CCGCTGCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-409_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCCTGCAACTTCAGGCCATT CAGATAGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-599_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTCTGGAACAGGCTGTCAG TACCGGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-529_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGACCATATTCAACAGGCC GCCAGGACT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1514_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCACCACGCCAGGTAAACT GTTTGTCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1253_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTCGTAACCAAACCAGGCGT TGATGTCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-490_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTAAAGCACGGAAACCGGGCC ACGCATAAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-176_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTAGCAGACTGTCGCGCCG CAATAGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1471_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTCCCGGCAACGCGGTTAAG AGATTCTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1799_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCTTTCGCTTCGATTTCTA CGCCACGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1183_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGATATCGCCAGTGGCAAA CTTGCTCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1100_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGTGCCAGATAATGGCCTT TATCCGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1029_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGGAGCAGACGCCGTAACCA TAAACGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1871_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTAGTATCGGTGGTGTATT CCGCATCGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1218_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACGCATAAAGTCGACACCG GTCAGCAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-264_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACTTCGGCTGATGCAGCGCC ATCTCTTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1323_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGGATCTTTGGCATTGAAG TCGAAATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2129_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGATCGAACAGGTTGTTAA CATGCAGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-873_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCATTAAAGTCTGTCGGCAGA CGCTTACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-455_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTCCAGCATATACGGGT CAATGACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1693_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACACGGCACCAGTAACTAC AATCGGACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1435_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTCATAACGACCGCCTAG GGTGACCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1600_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCATCCTTCCCAACTTGCGA AGAAGGTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-118_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGGTGCAGCGGTAACGGT GATAGTGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-939_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATTCGTGATCGAAGCTGTAG CCGACCATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2167_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAAGCAGCCATAAGTGTTAAA GCAGCTGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2029_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCCACTTTAAAGGAGTTAGC CGGATCACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2215_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAGAAACGGAAGGTTGCGGT TGCAACGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1288_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCGGATTGTACAGATTGAG CAGTGGCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-734_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAGGTGAACGCCGGTGCAA TAGCATAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-664_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAAGACCGGTCAGGCGATA AGAGTAAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1399_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCCCACTGCGCCTGATCCTG AACATAAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-2084_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATGCCGACTCGCGCCAGAT CATAACGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-309_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTACCAACAGAGACACCCGGC GTGTAGCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-358_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAAAGCCGCGAATGATCAG GTGGTCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-564_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAACTTCTTTCAGCGGTTCG GTGGTCGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-47_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTAACCGCTGTGGCTACTA CAACTGCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-82_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTTCAACCGCTGCCTGTGC ATAAACAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-796_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGGTTTCCGGCTCGTTCTG GAAGTAAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-699_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCTTCTGACCCTTTCTGCTGG GCATTGGCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-980_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCAAAGCGCAGGTTCTGAC GCACAGTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1764_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTCGCCACCTTCAACCGAG AAGAAGGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuA-1966_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGTCAGACCTGAAAGCGG ACCGTCAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-257_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGGAAGCTGCTGCGGCAAAT AAGCCACTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-536_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGACAGCACATCAACCTGGT GGGCGATAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-133_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGCAGAGTGGATTTACCAGA ACCGTTGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-720_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCATCGGGATGCCATAAATC ATTTCGAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-221_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACGCTTTGCTGCTCCAGC TTTCCAGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-639_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCGCAGGGCGACCAGATAA TCACAGTAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-310_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACGGGTAACGACCAATCGC CACCAGTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGAGGTCGGTTCGTCGAGCA ACAGACAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-95_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCGGTCACTTTCCCGGCAG GAAAGGTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-578_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATGACCGTCAGGCCACGCT CCTGACTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-45_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCGCCCGGGCACACGAAAG GAGATATTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-431_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGTTCGCCGCCAGAGAGAC TATCGACCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_fhuC-680_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATAATTTCCGCAGGCGTTC CCTGAGCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-321_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGTGCGGAACGGTCAGGCGT TCAGTATTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-16_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGAGAGGCCAGATTGCTGCC TATGGCAAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-281_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCAAACAGCATGATGTCGA GATCCAGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-391_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCAGGAAACACCAACTCCGG CGCGATTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGGCTGCGTTTAAGTAATC GGGTTGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-80_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTAAGAATGTGGCTTTCAG GGATATCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-183_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTAGCTCTTCAGGTGCAAGA GAGGTTTCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-356_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGCCACAGCATAAATCCAC GATTCTTCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_folK-230_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTGCGGACGCGACCTTGCT GCAATTCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-866_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCGATCACCTCTGCCAGTG TCTGACGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-608_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCGATATCGACGACGCAGA CACCCAGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-721_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCGTGCCAAAGGCGTAAGC GATATCACT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-43_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACTAAAGCGGCAACCTTCGC GGTACCAAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-429_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCACGCCCGAAAGTCCTACC GGATTCTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-475_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCATATCGTTGTGACATGT GATCAGGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-264_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCACCATACCAATTTCATTC TGGCAGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-117_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTTTATCCATACCACGCGAC GGGCAGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-393_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCTGATAGTCAATCGCATAC TCTTGCGGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-679_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAGCATAAGGAATTACCTT AGTGTGGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1136_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTCTTTCCCATAGTGAAGCA ATCCCACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1095_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGGCTCCTGAGCATAATCC GTTAAACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-82_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAATGATATTGACCATACC GTCGGGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-526_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTTGGTCAACTTTCAGCCC ACAACGTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-196_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGACAATCTGCCATCAATTC TGCCTGGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-907_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCTCTTCGTTGACCAGGTT GAGCAGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-768_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGCGCACAACCGTGGCGA ACTTTAATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-562_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATACCGAATAACTTGATGC CAGTCCGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1040_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCACTTGCGTATGAAACACGC GCTGAGCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-1221_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGCCAACTATTGAGTCGC TTGATCCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-998_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGATCTGCGCTGCGCCAC CGGTTAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsA-161_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTGTACGCACTTGACCA CGGATTCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-423_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTCTGGCGGCACGCTGAAG GTATTTCCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-464_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCCGGGCCATACAGCATTG GAAGCACCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTCGCGATAGCCCTGCAACA CTTCATTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-773_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTGCTGAGTAGATTCCTCTG GCGGCAAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-216_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAATGCCAGGATCGACTGC CGGATATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-290_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATCCACGGCAGGCGTTGTT CTATTTGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-251_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTTGACATCCTGGGTCATAA AGGTACCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-589_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGTTATTCAGCGTCAACTG CCAGGAACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-47_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGACGCGTTCCATTATTGC GGCGAGAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-554_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTCATCGCCGCTTCCTTCA GAGTAAATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-702_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTAGCTAATCCGTTTGCCA TCGGTTTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-388_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCTACCATATGTTGATCATT CCACCGCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-329_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAATTCATCAGGCCACTGCT TTCTGACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-738_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCCTACTGCCGCTCCAGAG TCATAACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-181_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTGTGTAATGGCGTTCACC GGTCAACAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsQ-12_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCTTCTTCGCTGTTTCGCGTG TTCAGAGCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-829_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCAAATGCACGGATGGTGTT ACCTACCGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-552_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCACAGCGCCTTTCAGTACA TCGTTCGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-864_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGAAGTACCGATAACCACA GTCGCGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-969_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCTGAACCTGCTTATTGGTC ACCAGAGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-899_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTTACGCGCAGCTCGTCAT TCATATCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-505_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGGAGATACCGCGGCCCAG AACTTTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-742_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGCCAGACAGGTCGATATC TTCCAGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-781_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGTCGAAGCCCGCCGTGAT GTTAACCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-934_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGACGTTTGTCCATGCCGAT ACCTGTCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-706_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGATAGCCATTTCAGCAGC TTCTTCCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-291_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCACCCATACCCGCAGCA ATAAAGACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-326_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCAGCGACGACTGGTGCTG CACCTGTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-182_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGCCCAGTCCTTTGGTGA TACCGCTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-1069_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGTTTGCGGCGCATTGTC ATTCACGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-622_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATTACGGTGCGTACGTCTGC AAAGTCCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-587_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAAACCCGGACGAGTAATCA GTTCAGCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-377_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAAAGGCTTAGTGACGACAG CAACGGTCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-27_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGACGCCGATGACTTTAATC ACCGCGTCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-1004_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCCCATGCTGCTGGTAGC GATCCATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-470_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTCGTTCGGGATAGTGATCA GAGAGTCCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-118_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGCTTGTGCATCGGTATT TACCGCGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ftsZ-77_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCAATGCGCTCGCGCACCA TGTGTTCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-965_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCCGGTTTGGTGTCTGGCA TATTGCTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-820_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCACTTCCGGCGAAGCGGT TTCTGCTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1189_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGTGCCCAGGAGTTTGGCGA GTTAAAGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-785_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGATAAGAGACACCACGGC AGGTTACGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1154_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTGTTCGTCAGACGGGATTG CGTTCGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1502_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGAACCACCAGTTCGCCAT TACGACGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-368_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCGCTAATCAGCAGTGCGA CCACCAGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1447_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCGGAGCGTAAATAACTGG CACTTTCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-644_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCCACTTTAATCGGCGTCA CTTCATTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1975_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCTGGCTGTTTACATGGCAG ACCAAATGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-877_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATCGCAATCAGTCGACCATC ATTGACCGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-930_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGAGCACGCCTTTATTGGCG AAGGTTTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1616_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACATATCCGCACCGCTCAAAT CTTTCGTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-542_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTATCGGCGTTAATTTGTT TCAGCGGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-23_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGGCTGTTAGCGTGACGAGTA ATCGTCGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2325_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGTCGCCCATCTTCGTACCA AATGAACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-126_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGAGCATCACAAGGCCAGCG ATAGGGTAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-403_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCGGATCGTTAAATCCGGC CCAGGTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-679_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGCGCTGGTGAGCGGTACA CAGGTAAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1729_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTACCCGGGAAGACCAGCGT ACCCTGTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1000_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGGTGATAATCGGTGGCGA AGTCGGTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1387_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTGCCGGAAGATCCATGTC CCACAGGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1240_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCACGCCCATCGGCAGATA GACCAGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-728_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGATCGTAATGCCATTTCT CTTTGCCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2362_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTACTTCACATCATCCGGCAG CGCATAAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1352_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAACGGTCTGGTAGCTCCACG CCAGTTTCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2290_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCTGCGGAGATCACCACATA CTGCTTACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2159_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCGTAGCGGCGATAAACA GCACGTTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2255_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCATAGGTCATTGGCGTAG CCTGACCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1789_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAATCGCCACTTCACGATT TGGATCAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGAGTCAGCGCCCAGAAGTC GAAACCAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1940_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAACGGGTTGAGCGTGACAC CATACGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2194_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGTTGCTCATGTTGTAAGC GCGCAGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-603_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGATCGTTCGGCTGCTTCACA TCGCCAGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-76_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCAGCCAGCCTCCGCCAAT GAGTAGATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1682_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCCTTCATAGCGCATCTGGT GGAACATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-507_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTGACCTTCCTGATTACGA CCATAGGCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2048_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCGGCGTACCAATACGTT TCTTCCACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1839_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGGACCACGCGGGATCAGT TTCGAAACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-2113_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCGCCCAGCATCGGCATACC CATATTGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-298_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACAAACGGCAGGATCAGCCA GATGCCGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_gcd-1066_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATCACGCCAGACGTTTCGCG GGTTGAGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-112_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGCCGGACCAGCTCTGACC TGAATGTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-407_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCTTTCGCAACGAACTGCA CGAAGTGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTGCCTACGGTGTCCATATT TGCGGCGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-458_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCAGTCACTACGTTACCAG CACAAATGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-351_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACGAAGTTTAATGCCGGG TTCAGGTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-760_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAATTTCTCGCCGTTCTCCTC AACGATGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-39_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGTGGAGCGTTTAGGGCGG ATGAGAACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-641_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGTGCAGCCACCATCGCTGA CGATCATTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-499_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACGATATCGGCACCTGAGAG GATAAGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-989_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGTTCCTGCACACGAATAA ACGTGGTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-74_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATTGACGTTCCAGTTCAACAT CGGAACGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-541_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGCGAGTTGTACAAACAGA ACCTGGGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-811_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAACGTGACGTTTCATCGC AGACTCGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-598_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAGCATCGGCACATTCGAT TACCGCAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_guaC-302_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAATCCGCATCAGACGTAC CGGTAGAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-416_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAGTCAGCGTGACCATGGT AACACCCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-298_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCATATCCATGGTCGGGAC CAGTTCGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-996_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTCTTCTGCCGCTTCCAGC AGACCTTCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-250_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACTTCCATTTCGGTTGGTGC ACCAAAGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-628_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAACTCTGGCAGCGGCGGAAC ACAGTTCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-961_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTGTTGTCAGCTCATCCAG CGTTTCGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1211_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATTGATATCTTCCATGCTGT GCGCCACGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-708_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTACGCGGAAACCGGTCATC ACTTCATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-120_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCCATCAACATCGTACAGA TAAGCGCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-762_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGAGGCAGGTTAAATCTGGC ACTACGCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-547_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCAAATGCGGCGCGTACAGA AGCCAGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1249_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAACTTCGCAAACACCCGACG TGCAGCATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1113_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTAAAGCGTTCCACGTCA CAGGCCATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-500_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTATATTTGGCGAAATCTG CCGGAACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-333_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCATGGTCGCTTCAGTGCCG GAGTTCACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-830_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCGCATCCATTACATCAC GACGACCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1171_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTCAAACGCTGACGGTGC CAGGTAAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-923_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGCGCGACTTCATTCAGAC AGGCGAAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-888_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCATCGCAATCGGGTTACCG GAAAGCGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-1040_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCGAACATGCCGCCAACGT GGTTAACGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-452_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTAACGTGAGTGCGCCAGAAC CGGCTTTCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hemL-593_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCACCGGCTCGACGATAA TACAGGCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-483_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGATATACGGCAGATGACGG TAACGCTGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-80_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATATCGCTGCCGCTGTCTT TGTAACGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-330_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTCGCGCAGGCTTAAGATC TCACGCACT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-153_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCATGAGATACCTGAACTTCA CGGCACAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-365_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTTATCCAGCAGCGTACAAA TCGCCAGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-400_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAATTCTACCGGGACGTTCAC TTCACGACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-271_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTTCAACAATCAGCACGTC CTTGCCACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-189_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCTACCGTAGCTGGAGGCG GTCATAAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-447_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAATGCCGTAACCCACCACA AACTCATCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-115_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAACATAAATGAGCCACGCAG CAGACCCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_hpt-41_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGACCCAGTTCGGCGATAC GCGCTTTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-376_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTATTCCCGGCTGTTAGCGC TTCAGATTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-339_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTCTTGCAAACGCGAGTTA AAGCGACTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-304_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAATAATACGCAGACGCAC GTTATGTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-180_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGGCATACAGCGTTAACGCC TCAATACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-411_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCGTAGTTCGCCGCAATA TTCAGCGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-486_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTGATCTGGTTGCAGGTTT CCTTGCTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-106_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCTTTATGCCCAAAGGCACG AATCTTCCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-446_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTTGCCTGACTCCCTGGA CTATATCCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-655_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCTTGTTCATCGAAATCGGG CCAGAGAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-257_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTATCGAGCGCCCACACAA ACAGTTCCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_ispU-706_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGCCGAAACGACGCTCTCG ATTAGCAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-389_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATGCGTAGCGCCAGGTACA GGCTTTCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-151_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAGGCAGGCAGAAACCGCGTG TTGTTCATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-301_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAGGTATCGCCGTGGAAACC ATCTTTGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-766_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTATTCGTCGTGCGAGATTAT CGCCGGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-79_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGCCCGGTTTAACATACGG TTCGATCAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-187_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGAGATGCAAACGGATTTCGG ATAGCCGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-114_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAATCATTACAGATGCGATCC AGCTCGCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGCCATCTTTCAGCAGCTT AGCATCGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-597_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTTGACCATTGGCTCGATG GTGAACGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-639_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGTCCAGCCATCTTTCATG GTGCGGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-731_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCTTGCGTAGCGTCAGAA TTTCGCAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCGCAATATTCACGAACGAC GGAGAAGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-516_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCGGTTCTTCATGGAAGCCG CGACCAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-562_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTTTCAGTACGACGTTGGT TTCACGGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGACAAATTTCTGAATCG CCGCACCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_map-338_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCCATGATGGTCGGCTTAC CGACGATAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1945_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCTGCGGGAAGCTCTGATA CAGCACTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-146_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTACGCCCTTTGCCTTTGC CCTTACCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1588_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCAATTGGCGCATCCGCAAT CCACGTATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1167_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCGGACGGGCACTCAACATG TCATAGAGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1001_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCTTAGCTCTTCTACCGGGC GACCAAAGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-861_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGATCAGCGCCATGTAAGCT TCGTTCGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1670_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCAGCATCACTCTGCCGC TTTCGCTAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-944_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGATTTCGTTGTCGCCGCTCT GACCGAGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2181_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGGCACCATACAGTTTGGTC GGCTGGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1036_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTCACCATACCGACTAACAG CGCCTGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-446_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGACGGCGAATCATCTCAA TGCTGTTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1635_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTCATCATTCTGCGGTGAC CAGACCTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1710_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGATTTACCGTCGGCACGTTC ATCGAACGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-325_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCATGTCTGGCTCAAGATT GACCATTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1457_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACTGCGGCTCAGAACCTC CGACCATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-896_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCAAGAATACGGTCTTTGC TGTAACGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2493_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACATATCCTTGATCCAACCG GCTACACCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-826_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTAAGAACGCTCGCTGGA GAGGAACAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-582_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGGTGATCAGACGCGGATCA AGACGGAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-617_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAAACAGACGCTGCTCACCGT TTGGCGAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-363_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGCCTCCAGCAGCTTCACC ATCTCGTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-19_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTCGGTTTCCCTTTGCGTCC AATTGGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2425_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCTGTTGCTGCGGCTGAGA AGACTGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-541_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGTTGTTCTCCATATTGAC GATCGTCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1543_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCTGGCTTAAGGCCGTCAG ATAAGTCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-234_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGAGATAAACGCCGTAAATG GCGATCAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1128_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATAATCTGTTGCTGTTGCAGC AGACGCAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-678_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGACGGTCTTCTGTCGCCAGC AAAGTATCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1377_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCTCAACTTACGCTGTTTC TTCAGTGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-784_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTTGCGTCAGCGTACTCGC ACCCTGTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2244_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAAGATTCAGCGGCGTTGGC GTCTGGTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-736_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTTGCCAGCACCGCACGTCC GATTGAGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2117_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCCGTCAATGCCCGCAAACC AGGTATCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2012_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTTGTACCACCTGCTGCA TGGTCCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-282_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGGCAGTTGCCAGACCTTG CCATCAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2286_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTCGTAGTCCACGCCCATA TCTGCAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1422_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACTTCACCACTAAAGCGGTCG ACGACCACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1766_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTACGCCCAGTTTAATCCAGG TCTCCGTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1241_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCAGCTCCTGACGCACCA GTTGCATAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-1315_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCCTGGGCCACCGAGTCAAA GGTAGTGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-481_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCACCTGTCCTTCTTTACT GTCCGGGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mrcB-2321_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAAGATACGCATGCCACCGC TGCAAACAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-253_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCGTCCGAGACAACGATATC GCCCACTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-323_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCGGACAGCCTGGTAACT GACCGTATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-366_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGCCTCAGCGGCAGCGATC AGTTTATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-67_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCGCAACCGCCGAGACTGAT AGTTTGACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-634_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCAGGCTGGACTGTTTAGC GGCAACAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-183_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCACATCTGGCTTGCAGTGT TCCAACAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-448_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCAGACCAACAGAACCGTT GATGAAAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-594_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGAAGCTAAGATGAGACTGT TGATCGGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-669_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGCCATGTGCAAGTTTCTGC ACCAGTGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-559_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAGATGGCGCGTACGACAAC AAACGGGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-102_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAACCTCGGTTCCATTCAGT TGGCCGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-512_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGGGCGATTGCCGTCGCTT CCATCTCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-137_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCAGCGACTTTACCGATGC CCGATTTCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-413_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTAACAATCAGGCCACGTA CAGCGTTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_mtn-288_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAAATGCCGTGACATCCGCG TCGTGATAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-1430_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTCCGGAGTTTGCGGCTTCA GTTTGATTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-282_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAATACGCGCTTCATGAGCG GCGACAATT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-1064_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTTAATGGTGGCGTCCACTT CCGTCGGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-698_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCGCGGTAACAATTCGCGGA TCACCGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-741_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGGCATCTTCGCTGAAGCCG TAAGTCGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-550_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTGACAATCGCCACCATCGG TTGCAGATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-597_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCTCAAAGTCGCCCTGGTAG GTATCCATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-402_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTCTGCGTAGATGCTGGAA ACCATCGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-911_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGCCTCGTCGTCAATGCCCT CTTCCGTAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-156_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCATTAACTGCTGCGTGACC GGATTTGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-106_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGATAACCTTCATTGGCCAG AACTTCGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-994_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCATTCACTGGCTCCAGCGG GAATTCACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-191_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGGCGATGGTTGAAATAAA TCGTCGCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-776_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGGCCAATCTGCTGATAAT CTTCTACAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-1268_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGTCCACGAATTGTGCGAC ACAGCGAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-247_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTATCGGCAGAAATCGCGCT GGAAACAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-1337_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTTAATACCGGTGCCAGCA TCTCGGCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-321_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCATTAACTCAGCCAGCATT TCGGCACGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-1214_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGCCAGCCGGATACACTT CCAGCATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-515_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGATGCATCACTCTCATCTG CTTCGGCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_murC-1179_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTATCAACCTGCGTCAGCACA TTGGCGAAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-616_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCACGAGGATCTGCCCGTC AGTGACGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-578_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATGCCAATAACCGGGATCG CCAGTGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-304_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCATGTTAGCACCGGCACG CATAACCGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-62_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGCGAAGCTATAGTCATAAG CGGTGATGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-653_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGGAATGTGACCGCCGGTAA TACCAAAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-8_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTACTTCTGCAGTAAGGAGA TGGTGGTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-762_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGGAAACTGTGTTCTTCGCCC GGATAAACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-490_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCAGCTTCTAAGGCTAATGC ATCGCTGAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-727_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCCACTTCAGCCATATACTG CCGCACAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-386_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACCTAAGTGACCACATACAG GAACGGCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-351_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCAGCATTTGTACGGTTTCT ACCAGCCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-180_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTGTGGTAGGCGATATCG GCAACGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-242_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATACGCCATAAACGGCAGGT CAGCCAGCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panB-113_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCGAATCGCCCACCAGCA TGACGTTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-775_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGGCGGCTACCAGAATTAC TGCCCGTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-406_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATGTCCGGCTGGACCAGGTT GAACAGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-37_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTGCCTTCCATACGCAGGCG GCGAATTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-226_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTTTGTTTAGCTTCTCGCA GTCCTCCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-297_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTAAGTGTGGGTTTCAGTA CCGTTCGGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-565_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTTGTTCCGCCGTCAGATA ACCGTTACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-76_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGTGCAGGTTACCCATGGT AGGCACCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-523_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGACCGTCTTTGGCGCGCAT AATTGGCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-111_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGGCTTTGGCTTCGTCGACC AGCTTCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-2_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCAGCGGCAGGGTTTCGA TAATTAACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panC-332_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACCTTCCAGCATGGTCGAAA GGCCAGGAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-26_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCATGAGTCACTTTCACGC GGTGGAGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-274_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGGTGCGAGCTTCTTCATC TGGCATGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-156_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGCGATGGCATAAGTGGAG AAACGCTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-63_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCCTGGTCAATGGCGCAAGAA CCTTCATAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-352_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAAGCAACCTGTACCGGAAT CGCTTTCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-239_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGCTGGCGATGATGACAATAT CGCCGACAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_panD-204_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGTGGGCCGCCGCACCGTTA ACAGAAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1007_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAAAGCGTCGTGATAGGTCA GGCCGCTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-652_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTTTCCGGGCTGATGCGCAT ACCCAATTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1102_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCAACTGCCAGATATCGCG GGTTAATGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-920_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAACGCCGGGTTCACGCGCA TATCGTTAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1314_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGGTGACGGTTCTTCATCC AGCTCGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-960_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGTCTCCAGCAGTGGGTAC CAGAACATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-478_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACAGGCTGTTGATAGTGAA ATCGCGGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-728_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTTAAGCGATTCTTCAAACA GGCGTGCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-767_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGCTTATAGGTTTCGTAAC CGTAGCCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-268_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTTACGGAACAGTTTGCG CACCTGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-569_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTCCGGGTTACCAATCAGAC GGATAACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1176_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCATAAGCCGCACGGAACTTA GGATGCTCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-533_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCAGATCCTTCATGCCGCCAA CGTAATCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-371_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGTCGCTGACGTTACCTTCGT GGTGTCCAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-336_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGCAACTTCGATAATCTCC GGGCCAAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-168_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAACCAGCCAGGCTTCGTAT CCCGCTTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-604_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCAGCATACGTACCGGATC TTCACGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1137_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGCGCGTTTACCCTGACGA CGGGACATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-885_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGTATCGGTATTCTTCAGC ACCTGTTCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-1211_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTACGCTCAACTTCAGCTC GCAAGGCCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pcnB-824_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTGAAGTAGCGGGTAATGG TCGGGAACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-497_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGATCAGCGGTAAACACGC CGTCAACTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-350_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAACAGGCTGATAGCTTCTG CCCAGCTGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-551_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGCACTTCGCTGTAAGTCA GTTGCTCGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-462_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTTCAGCACCACATCGGCT TCAATTTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-139_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAATCACCACACCAACCTG AATACCCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-91_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCCTGAGCCATACGATCCAG TATGCTTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-618_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAATCGGTAATTTATGGTCA CGAGCCAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-315_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCACACGCCATTCAATGGA ATAGCGGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-42_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCCCTGCAGAGCTTCGCCA CTCAACTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-425_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGGCAAGCTGCTGAGTCGG TGGTAAAGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_pyrH-390_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTACCTGTACCGGCGGAGAGG ATCACCACA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-723_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCGACTTTAACGATCCTG TCGCCTGCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-959_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCGACGATGGCGTTGAACG GCCCATACT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1068_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGGCGATAGAGATCGGCCCA CTGAGGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-565_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGTAAATCGAGCTTTACATC CCGCCGTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-64_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGGGCAACCCAGAAATGACC AAATTCATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-758_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATCACAAAGGTCACCCACT GCGTTAAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-245_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATTATTGAAGGCATGGTGGC GGAGTTCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-866_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCGCTTTACCATTACCCGGTT TACTCTCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-793_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAACGCTAAGGATTTACCCGG GTTATCCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-650_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTACAGGTTCAATTTGCG GCCCACGAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1138_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTTCACGCTAATAAGCGCAAG AAACGGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-172_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCCAACGGGATCAGGGCGAT AACATATTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-913_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCATCTGGCAAAGGAATGAC TTTCGGCTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-481_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACCAACTGCAAACGCACGGC ATCCCAATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-530_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCCAAATGGCGCTACTGTAA TGGTGGTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1286_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAACAGCACCAGCAGAATCG AGCCAATGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-1008_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCAGCATACTGACCGTCAGC TTCATCAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-25_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGTACACCCAGTGCAACGAT GAACGAAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-423_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTTCCGTACCTGGTGCAATT TGTGCTTCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-600_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCGGATCTTCTTTATCAGGC TCAAACGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-387_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCGAATTGGCTGCTATTTCG CCAACCACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-134_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTATCAGTTCGCCGCCAGA GCGCCTTAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-99_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTATTGAGAAACGCTCAACG CGAACACCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_rseP-299_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGAAGTTTGCAACCGGACCTG CGGCAATAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-907_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCGTTACGTGGTGCATCAG CATGATGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1073_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTCTGGATCTGCACACCTT CTTTCGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-77_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCTCCGGTTCCATGGCAT TGATGATGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1859_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTTGGCAATCGCTTTAGTCA CCCACGGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1760_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGCATCAGCGCATCTTCCA TCGACAGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2308_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGGTGCTCTTTCCACAGGGA GTCAAGCGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2184_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTCACGCAGCGTCTCTTCA TGCAGTTCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1703_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GACGACCAGAACGACCGCGCA ACTGGTTAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1535_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCTGCCAGCTACCACCGA GCACAATAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1666_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGGGATTCGTGACGCTCGGT ACCGATGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2631_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGGGCAAGGATCGTTACGT CCTACTTTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2078_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCCCACATTTCTTCCAGCG ACTGTGGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-610_l1-bc25mer_83004-20_l2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGAGTCCACTTCGTCCACCAG CGCATAGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-247_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATACCGCCGAGTAACTGAAC GTCGAAGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1348_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATGGAGATAGTACCCACCAG CACCGGCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2427_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCAACGACTCCAGCATCGCT GCAAACATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-681_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTATACATTTCCGAGCTGTCT TCTGCCGGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-351_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCACGCCTTTACCGGTTAGT GCGTTCAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGTTGTTCCAGCTCCTCAAC CTCTTCAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-540_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGTCGCGCAGGTAGTCAAAG CCGTATTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1121_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGACGGAAGTAGTTCTGGA AGGTGATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2148_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTATCCAGCCACTCGGCAATT GGCAAATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-575_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGCTGTACACGTTCTTCAG GGCTGAACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1435_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTCGTTGGCGTGGAATTT GGCGTTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-454_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATGCCCGGCAGGTTGATACC GACAGTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1267_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTTCAGTCATGTAGACCAG GTCCGGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-853_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCATCCATGATGCCCTCTTT CACCAGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-955_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACGATGTAGTCGACGTCACG GGTAAACAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-763_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGAGAAGTGGCCTTCGCCCTG GAAGGTTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1500_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACGACCCGCCATATTGGTC GCGATAGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2343_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGGTGGATACCCTGACGCAGA TAGTCCATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1907_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGTTGCTTACGAATGTCGA AGTTACGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2221_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTTACGCTGATACACTTCGAT GGACTGCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1970_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGTTCGTTACGCTGGGAGT AAATGGCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1821_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTTCGCCTGGCTTCATACCC AGTTTACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-505_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGTACCGTAAGTGATGTCAGC TGCGTAAGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1626_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCTTCCAGTACCGCATCG TGACGTACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1392_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATACCGGCTTTGGTCAGTTCG TTTGACACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2005_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGTTAATGGTTTCGCTCAC ATCGCTGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1231_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGAATCATTGGACGGTTGGT CGGAACAAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-282_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTACGCATTTCGGCGATGCAG CGTTCGTTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-212_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACCAAAGACGCGCTTACTTG CCTCACGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2535_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGCATTTGCGCTAAACGCTCG GCTTCCATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2273_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATGACGCCTTTCTCGAAGT GACGCATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-386_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CACGTTGCGCCAGGTAGTCGT TGACGGTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-2570_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCTGCAGAGTCGTCATCCT GATGGCTAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secA-1007_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCATGGTACGACCGGTGT GTTCGTCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secM-45_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCATCCCTAATAAGAGATGC GGCCAGAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secM-346_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGCGCGCTGAGCGTATCCAG TAATGCAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secM-174_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCGCTTCCAGCAAGGCCAAT TGACCAAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secM-81_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTTGCTGAGCGCAGGCAAA CCTAAACTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secM-209_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AATCAACGGAATAGTTCGAAT TCGGGCGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_secM-311_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCCTGAAGAGGCAAAGATT CTTCAGCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-215_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGAATAGTGACGCTGGCAC CCTGTGGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-407_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGAGAAATCACGCCGCAGG TAGAGACTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-689_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCTGTCGGTTAAGTCTTCCG GTTTGGTGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-724_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCTTTCCACAGCGCAGCGGT AATTTCCTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-759_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTGGCATATTGCGCCCGTAA TAAATCTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-593_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGACGTTCACATCCACCATGT CATACAGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-139_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACAGGTTTCTGACAGGATTTC GGTCAGACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-371_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGCGCGGAAGGTACATAAAC CGCCTTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-301_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCAAGATGGGCAACGACCGT TTCTGGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-250_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCGATGAGTTTCGGGTCAAC CGGTTCTTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-85_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATAAGCAATATAACCGTCGCG CTCTTCGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-442_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTCAAGCTGGTGGATCAGGTA ATTCAGCGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speD-477_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGCGCACGCGATAATCAATG GTTACGATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-622_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCAACGTCGCTGAAGTAATG GCTGAGTTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-367_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGATCGTCGTAGCTACCGGC GTTATGGTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-159_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TAGATAAACTCGTCGCGCTCG GTGGTTTGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-194_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGGCCAGTAGCGGAACATGGG TCATCATCT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-229_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGCCGATAATCAGCACATG TTTCGCGTG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-462_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGATAGGATCGGTGCAGTCG GAGATAATG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-729_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCGCCTGAATAATTTCGGTT GAGAGATGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-783_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCCGTATGGATTGCCGGATTG TAATAACGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-332_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATACTGACGGCAGAACGATA CGACACCCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-122_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CATCCAGCGCCATTACGCGAC CAAATGCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-406_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AACGAAATTGACGCCATCGTC GATCACCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-37_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCTACCGCAAAGTACTGCCC AAACTGGTC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-666_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCATGATACCGCCGTAATAG GTCGGGATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-542_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTGCGACGAAGATACCGCCAG GATTCAGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_speE-824_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAGCCAGTGCGTCTTGCAGAT ACTGAGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacC-248_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAATGCCGCTGAGATCGCGGT AGCTGTCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacC-80_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCGTTAAATCGGCCATAT CTTCGGCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacC-21_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CAGACGGCCATCAGGCTGCCG AATAACACT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacC-315_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGACGTAGGCAAGCAGGCTT AAGGAATCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacC-203_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTCGAAGGTGTCGTAATTAC TGAGGTCCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacC-149_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCAGTGCGCGACGAATTTGCC CGTTAGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-506_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGACTGGCGAATTAAATCCA GCACCATGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-117_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACTGACATTACGGAAGAAA TGCAGCGCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-715_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAGCAACAGGCCAGTTCGAA ATCCAGACG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-613_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTGCGGATAAAGCTGTGAATC GAGCGACAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-210_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACGCCAATCCAGGTCTGGAGT TTACGTTGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-245_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAATGCTTCAATACGGCTCT GGTCCACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-385_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGTAGGTAAATCAAAGCTGCA ACAGCCGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-175_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GTCAAGTTCTTTCAACAGCTC AGTGCGGAC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-305_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GAAATTGCCCGATACGCGGCG CGGAAATTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-461_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGGGTTCAGGCTGGCAATCC AGGTTTCTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacF-349_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCTCAGTCGCTGACGCACCAG AGCAATCAA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacG-10_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCCGCAGGTTGGGCAATTCAC CGTAATAGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacG-169_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACTGCTTTGGTTCTTCGCT CCAGTCATC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacG-101_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTTCAGCAGCCCATTCTCCGA GGTCGATCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-284_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTCAACGCCGCATAGCGACA GGCTTTCTT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-191_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCATCACCTCTTCACCGTCCA TCCACAGGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-110_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGCTGCTTGTTCCACTTCAT CAAGCAAGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-145_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTTGCCAGGACCGTTCGCT ACCTTTCAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-226_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCATCGCCAGCGAATTCCAG TTGATTGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-333_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACTTCTGCTGCACGAAATTG CGGTAAGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-69_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTCACCTCTTCATTAAACCAG TGCCCGACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yacL-27_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATGGACATACGCACCTTTACC ACTCCGGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-726_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AGTCGATACTGATAGCCATCG AGCTTCGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-761_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCAGCACTTCAACTTCCAGCG TCGCGGTAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-575_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATTTCTGCTTCTTCCAGGT AGTGTGTGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-870_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACCAGTGAAACAAACAGCTCT TCCAGACGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-281_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCCTGATTCACCACAATTT GCTGCACGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-393_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ATCCCGCCAGATAACATACGC GCACGTTCG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-227_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCTGCGGCACCAGTCCCAACT GACGTTTAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-428_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCATTAACGCACGGGCAATCA TTAAACGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-497_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC ACATTGAGCGGCGAAGTTCAA TATCCACGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-182_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCTCGAGATCGTAACCAAATA CGCTGACCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-316_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GATGTACGCTTCTTTGCGCTC CACGCCGTA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-685_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGGTGCGAGATCGAGAATAAA GGTTTCCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-617_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCACCAGCTCACCGTGTTGAA TAATGCCGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadG-66_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TCACCCGCTTCGACCTGCAAA TCTATCCCA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-217_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGATGACGCAACGTTGGCGTA GGCATTGGT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-742_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TTAGCTACGCAAACCACGTCC ACGTTGGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-53_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GCGTCTGCACCCAGATACGCA TAAAGCGAT
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-88_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC AAAGTAAAGGGTCATGGTGAT GACTGGCGG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-463_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC GGCAAACACACCGTTCAGCAA ACCCGCAAG
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-519_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TACGTGAGTGGCGTTAACACA AAGGTTGGC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-151_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CTGCATATAGCTGAAGCCATG CATATCGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-599_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATAAACGATTGGGTTCAGGT GCGACAGCC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-297_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TATCCGGCAATAATGACGTGA GTCGGAACC
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-419_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CCGTGAGCACCAGCGTTAAGG CAACGAATA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-653_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC TGACCAGCGGAACATCATTGA TACCGAGGA
CATTGC GCTTCTCCGGAAACATTAGC

Exp-1_yadH-688_I1-bc25mer_83004-20_I2-bc25mer_83517-20
ACAACCGCGTGTTACAAGGC CAGGCATCCGAGAGGTCTGG
GAATGC CGCCACAATAAAGACCACCAG TACGCCAAA
CATTGC GCTTCTCCGGAAACATTAGC

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within

the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes; exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary

probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid

probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown

where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention; and

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:
GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGATC
ATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCGGACA
AACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG; and

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary and secondary probes are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target nucleic acid, and enough of the

target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then

fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-high resolution imaging techniques (e.g., resolutions better than the wavelength of visible light) may be used. Examples of super-high resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-high resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read

sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For

example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-high resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be

determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the cell or other sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword," e.g., using only first probe 21 and second probe 22 the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probe 21 while the second position corresponds to the binding of probe 22. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction,

double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may by short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein. Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to

produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least

75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid

sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30,

20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a

population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. One of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; fewer than this number of different probes may also be used in

certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, up to $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. Up to 2^n different probes may be possible for n rounds of probes, but the number of nucleic acid probes that are actually used may be any number less than 2^n. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid probe codewords. If a match is found, then the nucleic acid probe can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus the correct nucleic acid probe. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one of the nucleic acid probes is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one of the nucleic acid probes is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one

set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions.

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated

Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be

determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633

nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. . Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a

stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO

633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low

calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red,

Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding, error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular

(having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft lithography techniques suitable for this embodiment are discussed in the references entitled "Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified

from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers, and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the

oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective smFISH approach, but also allows researchers to benefit

from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in abundance or localization. Simultaneous measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a

different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the amount of error correction desired determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by

hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. Even for a modest number of moderately expressed mRNAs, the density of labeled mRNAs exceed that at which each mRNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of secondary hybridization and imaging, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from

individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps (meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be

included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging of super-resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended

Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into

two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 °C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10

B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to

synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-

microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters

may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0]. Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are

meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also

including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but

not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

CLAIMS

1. A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2. The method of claim 1, comprising exposing the sample to at least 5 different nucleic acid probes.

3. The method of any one of claims 1 or 2, comprising exposing the sample to at least 10 different nucleic acid probes.

4. The method of any one of claims 1-3, comprising exposing the sample to at least 100 different nucleic acid probes.

5. The method of any one of claims 1-4, comprising exposing the sample to a plurality of nucleic acid probes simultaneously.

6. The method of any one of claims 1-5, comprising exposing the sample to a plurality of nucleic acid probes sequentially.

7. The method of claims 1-6, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8. The method of claim 7, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of RNA species in the sample.

9. The method of claim 7 or 8, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of DNA sequences in the sample.

10. The method of any one of claims 1-9, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11. The method of claim 10, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

12. The method of any one of claims 10 or 11, wherein the target sequence is substantially complementary to a nucleic acid sequence encoding a protein.

13. The method of any one of claims 10-12, wherein the target sequence is substantially complementary to messenger RNA (mRNA).

14. The method of any one of claims 10-13, wherein the plurality of nucleic acid probes comprises at least 8 possible read sequences.

15. The method of any one of claims 10-14, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

16. The method of any one of claims 10-15, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

17. The method of any one of claims 10-16, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

18. The method of any one of claims 10-17, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19. The method of any one of claims 10-18, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

20. The method of any one of claims 10-19, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

21. The method of any one of claims 10-20, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

22. The method of any one of claims 10-21, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

23. The method of any one of claims 10-22, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

24. The method of any one of claims 10-23, wherein at least some of the plurality of nucleic acid probes comprise no more than 3 read sequences.

25. The method of any one of claims 10-24, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

26. The method of claims 1-25, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27. The method of claim 26, wherein the first signaling entity is fluorescent.

28. The method of any one of claims 26 or 27, wherein the first signaling entity is a protein.

29. The method of any one of claims 26-28, wherein the first signaling entity is a dye.

30. The method of any one of claims 26-29, wherein the first signaling entity is a nanoparticle.

31. The method of any one of claims 26-30, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32. The method of claim 31, wherein the first signaling entity and the second signaling entity are identical.

33. The method of claim 31, wherein the first signaling entity and the second signaling entity are not identical.

34. The method of any one of claims 31-33, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35. The method of claim 34, comprising inactivating the first signaling entity by photobleaching at least some of the first signaling entity.

36. The method of any one of claims 34 or 35, comprising inactivating the first signaling entity by chemically bleaching at least some of the first signaling entity.

37. The method of any one of claims 34-36, comprising inactivating the first signaling entity by exposing the first signaling entity to a reactant able to alter the structure of the signaling entity.

38. The method of any one of claims 34-37, comprising inactivating the first signaling entity by removing at least some of the first signaling entity.

39. The method of any one of claims 34-38, comprising inactivating the first signaling entity by dissociating the first signaling entity from the first secondary probe.

40. The method of any one of claims 34-39, comprising inactivating the first signaling entity by dissociating the first secondary probe that contains the first signaling entity from the sample.

41. The method of any one of claims 34-40, comprising inactivating the first signaling entity by exposing the signaling entity or the first secondary probe to an enzyme.

42. The method of claims 1-41, wherein at least some of the plurality of nucleic acid probes comprise DNA.

43. The method of any one of claims 1-42, wherein at least some of the plurality of nucleic acid probes comprise RNA.

44. The method of any one of claims 1-43, wherein at least some of the plurality of nucleic acid probes comprise PNA.

45. The method of any one of claims 1-44, wherein at least some of the plurality of nucleic acid probes comprise LNA.

46. The method of any one of claims 1-45, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

47. The method of any one of claims 1-46, wherein at least some of the plurality of nucleic acids probes are configured to bind to a nucleic acid within the sample.

48. The method of any one of claims 1-47, wherein at least some of the binding of the nucleic acid probes to a target within the sample is specific binding.

49. The method of any one of claims 1-48, wherein at least some of the binding of the nucleic acid probes to a target within the sample is via Watson-Crick base pairing.

50. The method of any one of claims 1-49, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

51. The method of any one of claims 1-50, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

52. The method of any one of claims 1-51, wherein at least some of the plurality of nucleic acid probes are configured to bind to mRNA.

53. The method of any one of claims 1-52, wherein at least some of the plurality of nucleic acid probes are configured to bind to transfer RNA (tRNA).

54. The method of any one of claims 1-53, wherein at least some of the plurality of nucleic acid probes are configured to bind to ribosomal RNA (rRNA).

55. The method of any one of claims 1-54, wherein at least some of the plurality of nucleic acid probes are configured to bind to lncRNA.

56. The method of any one of claims 1-55, wherein at least some of the plurality of nucleic acid probes are configured to bind to snoRNA.

57. The method of any one of claims 1-56, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

58. The method of any one of claims 1-57, wherein the plurality of nucleic acids probes is configured to bind to DNA.

59. The method of any one of claims 1-58, wherein at least some of the plurality of nucleic acid probes are configured to bind to genomic DNA.

60. The method of any one of claims 1-59, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

61. The method of any one of claims 1-60, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm.

62. The method of any one of claims 1-61, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 80 nm.

63. The method of any one of claims 1-62, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 50 nm.

64. The method of any one of claims 1-63, wherein the sample comprises a cell.

65. The method of claim 64, wherein the cell is a human cell.

66. The method of any one of claims 64 or 65, wherein the cell is fixed.

67. The method of any one of claims 1-66, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

68. The method of any one of claims 1-67, comprising determining binding of the nucleic acid probes using an optical imaging technique.

69. The method of any one of claims 1-68, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

70. The method of any one of claims 1-69, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

71. The method of claim 70, comprising determining binding of the nucleic acid probes using stochastic optical reconstruction microscopy (STORM).

72. The method of any one of claims 70 or 71, comprising determining binding of the nucleic acid probes using photoactivated localization microscopy (PALM) or fluorescence photoactivation localization microscopy (FPALM).

73. The method of any one of claims 70-72, comprising determining binding of the nucleic acid probes using stimulated emission depletion microscopy (STED).

74. The method of any one of claims 70-73, comprising determining binding of the nucleic acid probes using structured illumination microscopy (SIM).

75. The method of any one of claims 70-74, comprising determining binding of the nucleic acid probes using reversible saturated optically linear fluorescence transition (RESOLFT) microscopy.

76. A method, comprising:

exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and

for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

77. The method of claim 76, further comprising:

creating codewords based on determination of the read sequences within the sample.

78. The method of claim 77, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

79. A method, comprising:

exposing a sample to a plurality of primary nucleic acid probes;

exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample;

creating codewords based on fluorescence of the secondary nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

80. A method, comprising:

exposing a plurality of primary nucleic acid probes to a sample; and

exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample, wherein at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

81. The method of claim 80, further comprising:

creating codewords based on fluorescence of the secondary nucleic acid probes.

82. The method of claim 81, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

83. A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

84. The composition of claim 83, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

85. The composition of claim 84, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 3.

86. The composition of any one of claims 84 or 85, wherein the code is a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, or a Hamming(127, 120) code.

87. The composition of any one of claims 83-86, wherein the code is a SECDED code.

88. The composition of claim 87, wherein the code is a SECDED(8,4) code.

89. The composition of claim 87, wherein the code is a SECDED(16,4) code.

90. The composition of claim 87, wherein the code is a SECDED(16, 11) code, a SECDED (22, 16) code, a SECDED (39, 32) code, or a SECDED (72, 64) code.

91. The composition of any one of claims 87-90, wherein only codewords with a constant number of ones are used.

92. The composition of any one of claims 83-91, wherein the plurality of nucleic acid probes comprises no more than 100 possible read sequences.

93. The composition of any one of claims 83-92, wherein the plurality of nucleic acid probes comprises no more than 64 possible read sequences.

94. The composition of any one of claims 83-93, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

95. The composition of any one of claims 83-94, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

96. The composition of any one of claims 83-95, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

97. The composition of any one of claims 83-96, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

98. The composition of any one of claims 83-97, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

99. The composition of any one of claims 83-98, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

100. The composition of any one of claims 83-99, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

101. The composition of any one of claims 83-100, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

102. The composition of any one of claims 83-101, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

103. The composition of any one of claims 83-102, wherein at least some of the plurality of nucleic acid probes comprise DNA.

104. The composition of any one of claims 83-103, wherein at least some of the plurality of nucleic acid probes comprise RNA.

105. The composition of any one of claims 83-104, wherein at least some of the plurality of nucleic acid probes comprise PNA.

106. The composition of any one of claims 83-105, wherein at least some of the plurality of nucleic acid probes comprise LNA.

107. The composition of any one of claims 83-106, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

108. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position

within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

109. The method of claim 108, wherein the code space has a Hamming distance of at least 2.

110. The method of any one of claims 108 or 109, wherein the plurality of targets comprise RNA.

111. A method, comprising:

exposing nucleic acid probes formed using the method of any one of claims 108-110 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets within the sample.

112. The method of claim 111, wherein at least some of the plurality of secondary nucleic probes is able to bind to a read sequence of a nucleic acid probe.

113. The method of claim 112, wherein at least some of the plurality of secondary nucleic probes is able to specifically bind to only one read sequence.

114. The method of claim 111-113, wherein if no match is found between a codeword and the codeword derived from the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

115. The method of claim 111-114, wherein at least some of the secondary nucleic acid probes each comprise a signaling entity.

116. The method of claim 115, comprising imaging the signaling entity to determine binding of the plurality of secondary nucleic probes within the sample.

117. The method of claim 116, comprising imaging the signaling entity using a super-resolution imaging technique.

118. The method of any one of claims 116 or 117, further comprising applying drift correction to the images.

119. The method of any one of claims116-118, further comprising applying physical registration to the images.

120. The method of claim 111-119, comprising determining binding of the plurality of secondary nucleic probes within the sample at a resolution better than 100 nm.

121. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

122. The method of claim 121, where in the codewords comprise a number of positions that is less than the number of targets.

123. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a

plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

124. The method of claim 123, wherein the codewords form an error-checking space.

125. The method of any one of claims 123 or 124, wherein the codewords form an error-correcting code space.

126. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

127. The method of any one of claims 123-126, wherein the groups are applied to the sample sequentially using fluidic devices.

128. The method of any one of claims 123-127, wherein each group of nucleic acid probes contains a group identification sequence distinguishable from the group identification sequence of the other groups.

129. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code;

associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample;

forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences;

forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences;

exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample;

exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

130. The method of claim 129, wherein the wherein the read sequences are taken from a set of orthogonal sequences, which have a homology of less than 15 basepairs with one another and with the nucleic acid species in a sample.

131. The method of any one of claims 129 or 130, comprising combining the images of the sample into a combined image.

132. The method of claim 131, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

133. The method of claim 132, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

134. The method of claim 133, comprising matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword that correspond to a nucleic acid target

135. The method of any one of claims 129-134, comprising imaging the sample at a resolution better than 300 nm.

136. The method of any one of claims 129-135, comprising imaging the sample at a resolution better than 100 nm.

137. The method of any one of claims 129-136, comprising imaging the sample using a super-resolution fluorescence imaging technique.

138. The method of any one of claims 129-137, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

139. The method of any one of claims 129-138, where in the codewords comprise a number of positions that is less than the number of targets.

140. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes comprising a signaling entity and a

target sequence that uniquely binds to one of the nucleic acid targets;

grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords;

exposing a sample to one of the probe pools;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

141. The method of claim 140, comprising combining the images of the sample into a combined image.

142. The method of claim 141, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

143. The method of claim 142, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

144. The method of claim 143, comprising matching the determined codeword to a valid codeword wherein, if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target.

145. The method of any one of claims 140-144, comprising imaging the sample at a resolution better than 300 nm.

146. The method of any one of claims 140-145, comprising imaging the sample at a resolution better than 100 nm.

147. The method of any one of claims 140-146, comprising imaging the sample using a super-resolution fluorescence imaging technique.

148. The method of any one of claims 140-147, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

5

149. The method of any one of claims 140-148, wherein the codewords comprise a number of positions that is less than the number of targets.

ABSTRACT

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Encoding Scheme

RNA target	codeword	probe sequence
mRNA I	110	
mRNA II	011	
mRNA III	101	

Primary sites (RNA tag)
Secondary site 1
Secondary site 2
Secondary site 3

Fig. 1A

b) **Code-Reading Scheme 1: Primaries and Secondaries**

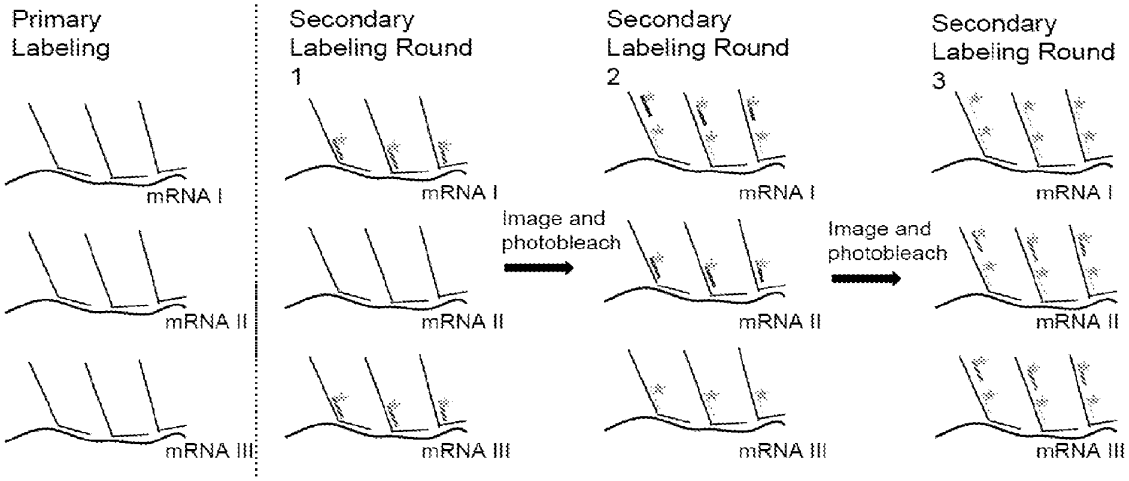


Fig. 1B

Code-Reading Scheme 2: Primaries Only

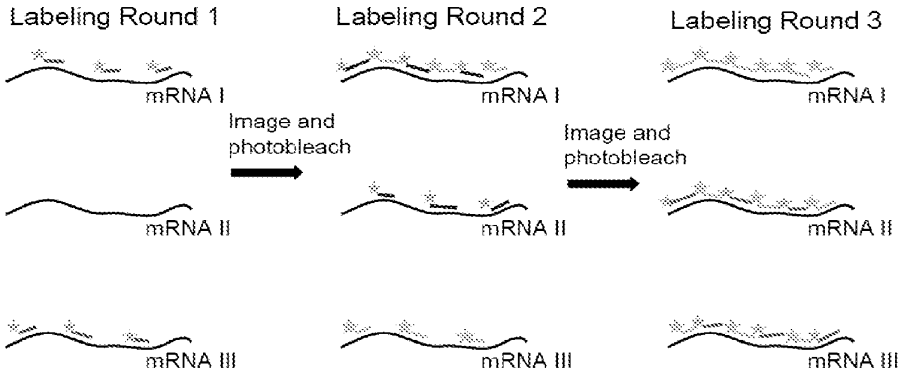


Fig. 1C

Fig. 2A

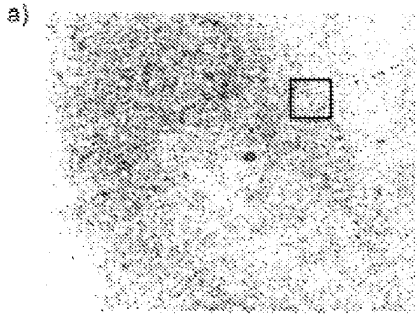


Fig. 2B

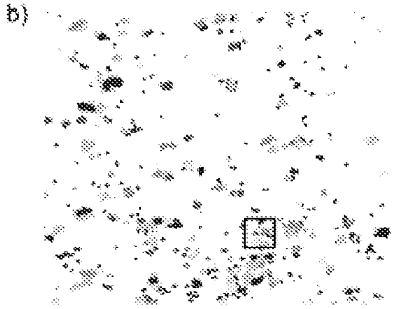
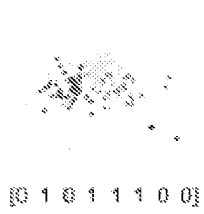


Fig. 2C



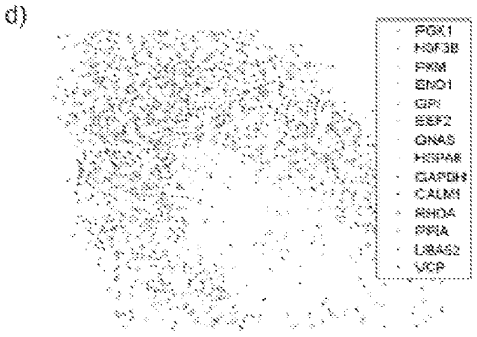


Fig.2D

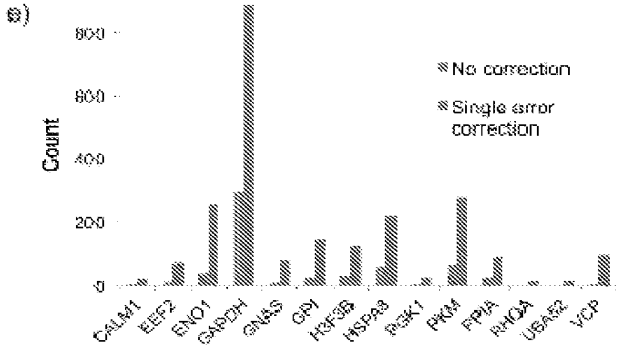


Fig. 2E

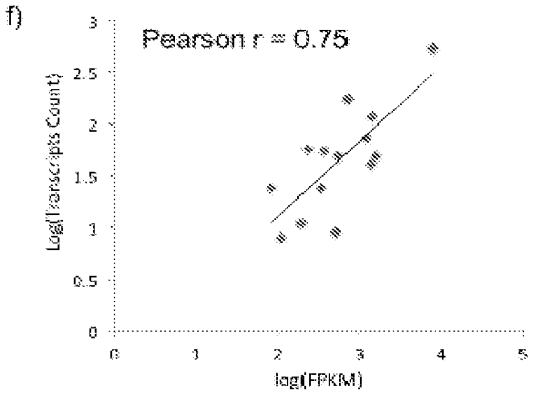


Fig. 2F

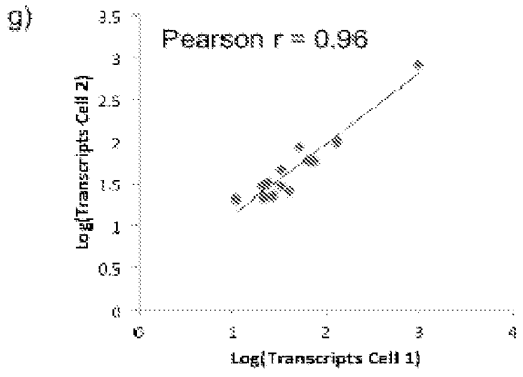


Fig. 2G

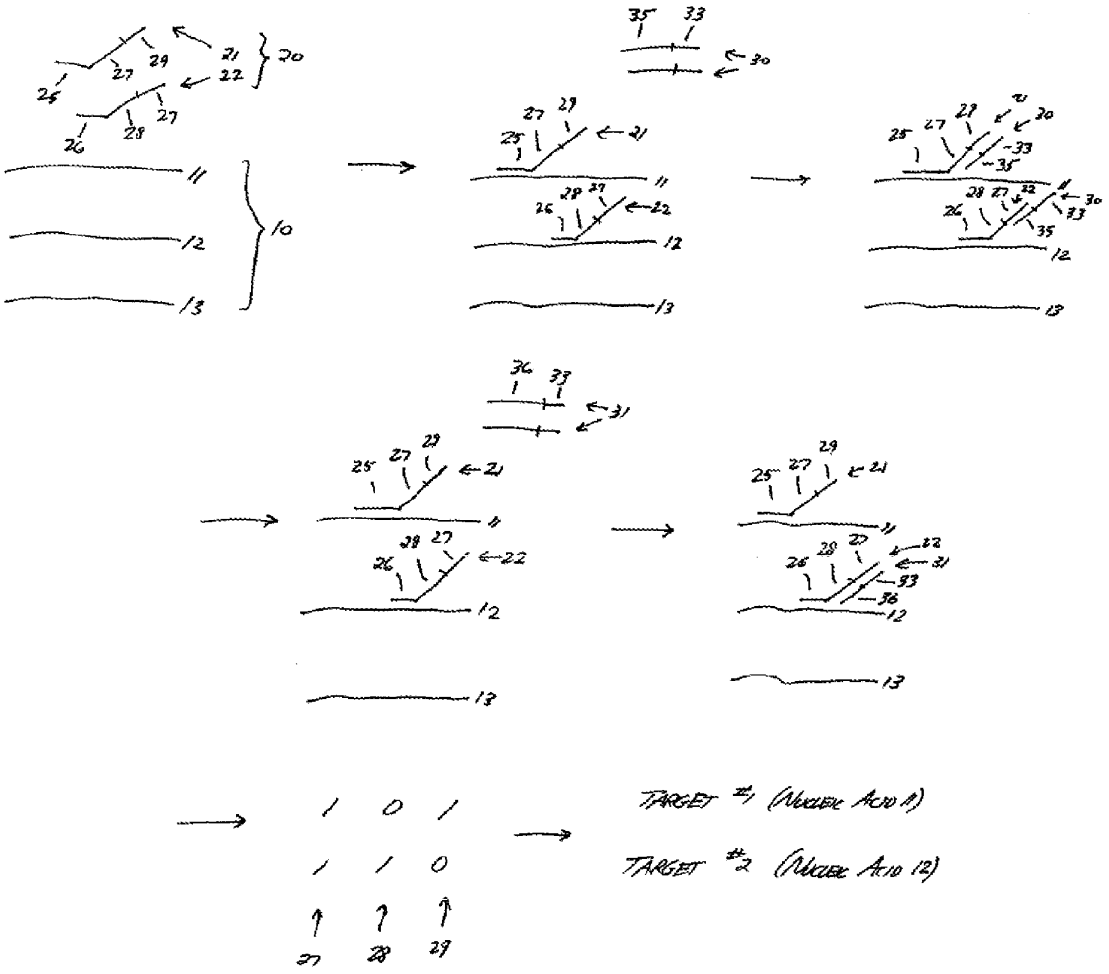

1 0 1
1 1 0
27 28 29
TARGET #1 (NUCLEIC ACID 11)
TARGET #2 (NUCLEIC ACID 12)

Fig. 3A

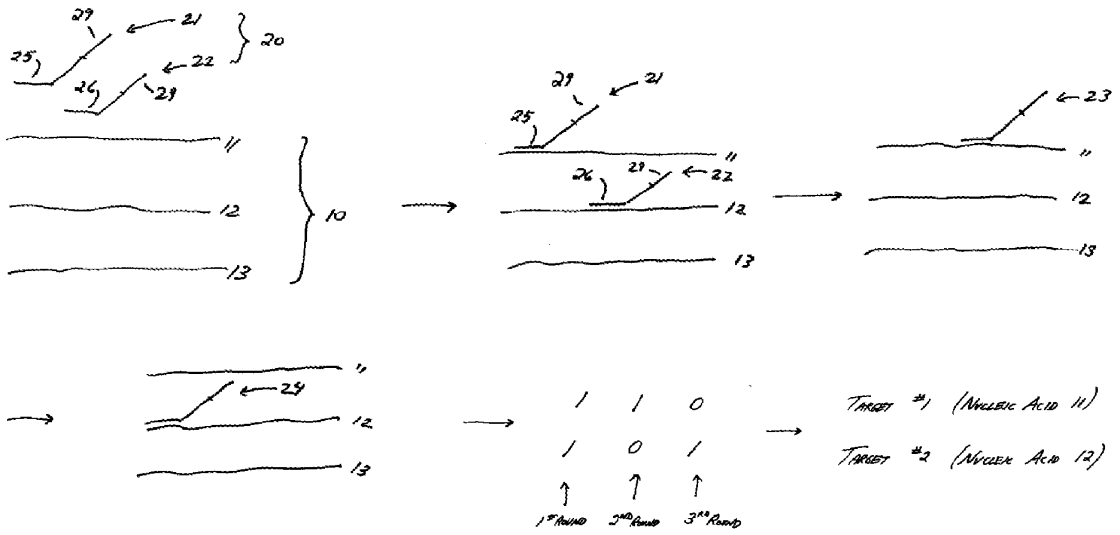


Fig. 3B

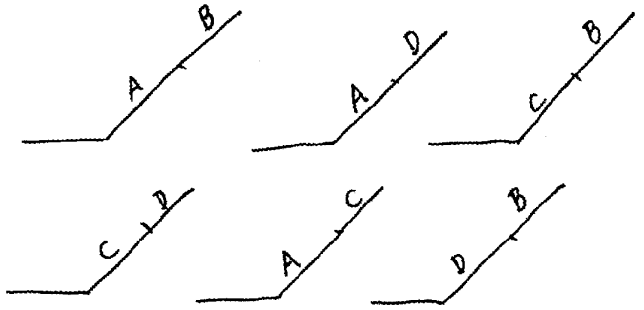


Fig. 4A

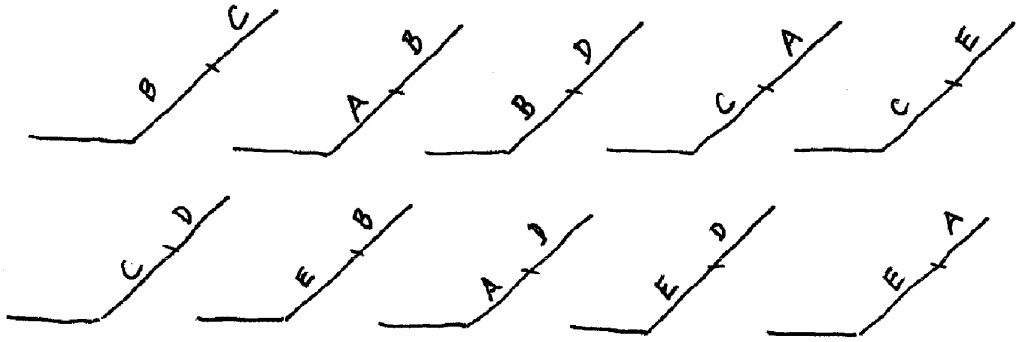


Fig. 4B

CLAIMS

1. A method, comprising:

amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides;

transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA;

reverse transcribing the RNA to produce transcribed DNA; and

selectively degrading the RNA relative to the transcribed DNA.

2. The method of claim 1, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides.

3. The method of any one of claims 1 or 2, wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences.

4. The method of any one of claims 1-3, wherein the plurality of oligonucleotides includes at least 100 unique oligonucleotide sequences.

5. The method of any one of claims 1-4, wherein the plurality of oligonucleotides includes at least 1,000 unique oligonucleotide sequences.

6. The method of any one of claims 1-5, wherein the plurality of oligonucleotides includes at least 10,000 unique oligonucleotide sequences.

7. The method of any one of claims 1-6, wherein the plurality of oligonucleotides includes at least 100,000 unique oligonucleotide sequences.

8. The method of any one of claims 1-7, wherein the unique oligonucleotide sequences each comprise a first region that is identical in the unique oligonucleotide sequences, and a second region that is not identical in the unique oligonucleotide sequences.

9. The method of claim 8, wherein the unique oligonucleotide sequences each comprise a third region that is identical in the unique oligonucleotide sequences, wherein the second region separates the first region and the third region.

10. The method of any one of claims 1-9, wherein amplifying at least some of the plurality of oligonucleotides comprises exposing a least some of the plurality of oligonucleotides to primer-containing sequences.

11. The method of claim 10, wherein at least some of the primer-containing sequences further comprises a promoter, whereby the promoter is added to at least some of the amplified oligonucleotides.

12. The method of any one of claims 1-10, wherein at least some of the amplified oligonucleotides contain a promoter.

13. The method of any one of claims 11 or 12, wherein the promoter is a T7 promoter.

14. The method of any one of claims 11 or 12, wherein the promoter is a T3 promoter.

15. The method of any one of claims 11 or 12, wherein the promoter is a SP6 promoter.

16. The method of any one of claims 1-15, wherein at least some of the plurality of oligonucleotides comprise at least a first set of oligonucleotides having a first common index region, and a second set of oligonucleotides having a second common index region distinguishable from the first common index region.

17. The method of claim 16, comprising amplifying the first set of oligonucleotides but not the second set of oligonucleotides.

18. The method of any one of claims 16 or 17, wherein the plurality of oligonucleotides comprises at least 2 sets of oligonucleotides having distinguishable common index regions.

19. The method of any one of claims 16-18, wherein the plurality of oligonucleotides comprises at least 4 sets of oligonucleotides having distinguishable common index regions.

20. The method of any one of claims 16-19, wherein the plurality of oligonucleotides comprises at least 10 sets of oligonucleotides having distinguishable common index regions.

21. The method of any one of claims 16-20, wherein the plurality of oligonucleotides comprises at least 20 sets of oligonucleotides having distinguishable common index regions.

22. The method of any one of claims 16-21, wherein the plurality of oligonucleotides comprises at least 96 sets of oligonucleotides having distinguishable common index regions.

23. The method of any one of claims 16-22, wherein the plurality of oligonucleotides comprises at least 100 sets of oligonucleotides having distinguishable common index regions.

24. The method of any one of claims 16-23, wherein the plurality of oligonucleotides comprises at least 192 sets of oligonucleotides having distinguishable common index regions.

25. The method of any one of claims 1-24, wherein amplifying at least some of the plurality of oligonucleotides comprises producing amplified oligonucleotides

comprising one oligonucleotide of the plurality of oligonucleotides and a promoter.

26. The method of any one of claims 1-27, wherein transcribing at least some of the amplified oligonucleotides comprises a mass of RNA that is at least 100-fold greater than the mass of amplified oligonucleotides.

27. The method of any one of claims 1-26, wherein transcribing at least some of the amplified oligonucleotides comprises exposing the amplified oligonucleotides to an RNA polymerase.

28. The method of claim 27, wherein the RNA polymerase is a T7 RNA polymerase.

29. The method of claim 27, wherein the RNA polymerase is a T3 RNA polymerase.

30. The method of claim 27, wherein the RNA polymerase is a SP6 RNA polymerase.

31. The method of claim 27, wherein the RNA polymerase is RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, or RNA polymerase V.

32. The method of claim 27, wherein the RNA polymerase is a viral RNA polymerase.

33. The method of any one of claims 1-32, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10 RNA copies of each of the amplified oligonucleotides.

34. The method of any one of claims 1-33, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 100 RNA copies of each of the amplified oligonucleotides.

35. The method of any one of claims 1-34, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 1,000 RNA copies of each of the amplified oligonucleotides.

36. The method of any one of claims 1-35, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10,000 RNA copies of each of the amplified oligonucleotides.

37. The method of any one of claims 1-36, wherein reverse transcribing the RNA comprises exposing the RNA to a reverse transcriptase.

38. The method of claim 37, wherein the reverse transcriptase is a viral reverse transcriptase.

39. The method of any one of claims 1-38, wherein reverse transcribing the RNA to produce transcribed DNA occurs without first purifying the RNA from components used to produce the RNA.

40. The method of any one of claims 1-39, further comprising purifying the RNA from components used to produce the RNA prior to reverse transcribing the RNA to produce transcribed DNA.

41. The method of any one of claims 1-40, wherein reverse transcribing the RNA to produce transcribed DNA comprises reverse transcribing the RNA to produce transcribed DNA using a sequence containing a transcription primer.

42. The method of claim 41, wherein the sequence containing a transcription primer is incorporated into the transcribed DNA.

43. The method of any one of claims 41 or 42, wherein the sequence containing a transcription primer further comprises a biotin moiety.

44. The method of any one of claims 41-43, wherein the sequence containing a transcription primer further comprises a digoxigenin moiety.

45. The method of any one of claims 41-44, wherein the sequence containing a transcription primer further comprises a fluorophore.

46. The method of any one of claims 41-45, wherein the sequence containing a transcription primer further comprises a non-nucleic acid moiety.

47. The method of any one of claims 41-46, wherein the sequence containing a transcription primer further comprises a signaling entity attached thereto.

48. The method of claim 47, wherein the signaling entity is fluorescent.

49. The method of any one of claims 47 or 48, wherein the signaling entity is photoswitchable between a first state able to emit light at a first wavelength, and a second state not able to emit light at the first wavelength.

50. The method of any one of claims 47-49, wherein the signaling entity is photactivatable.

51. The method of claim 47, wherein the signaling entity is a nucleic acid sequence.

52. The method of claim 51, wherein the additional nucleic acid sequence is able to hybridize selectively to a second nucleic acid sequence containing an additional signaling entity via Watson-Crick base-pairing.

53. The method of any one of claims 1-52, wherein selectively degrading the RNA relative to the transcribed DNA comprises chemically reducing the RNA.

54. The method of any one of claims 1-53, wherein selectively degrading the RNA relative to the transcribed DNA comprises exposing the RNA to a pH of at least about 9.

55. The method of any one of claims 1-54, wherein selectively degrading the RNA relative to the transcribed DNA comprises subjecting the RNA to alkaline hydrolysis.

56. The method of any one of claims 1-55, wherein selectively degrading the RNA relative to the transcribed DNA comprises enzymatically digesting the RNA.

57. The method of any one of claims 1-56, wherein the transcribed DNA is substantially single-stranded.

58. The method of any one of claims 1-57, wherein the steps are performed in the order recited.

59. The method of any one of claims 1-58, further comprising purifying the transcribed DNA.

60. The method of claim 59, comprising purifying the transcribed DNA by column purification.

61. The method of any one of claims 59 or 60, comprising purifying the transcribed DNA by ethanol precipitation.

62. The method of any one of claims 59-61, comprising purifying the transcribed DNA by solid-phase reversible immobilization.

63. The method of any one of claims 1-62, further comprising concentrating the transcribed DNA.

64. The method of claim 63, comprising concentrating the transcribed DNA via evaporation.

65. The method of any one of claims 1-64, wherein the plurality of oligonucleotides is computationally designed.

66. The method of any one of claims 1-65, wherein the plurality of oligonucleotides is synthetically produced.

67. The method of any one of claims 1-66, wherein the plurality of oligonucleotides arises from a cell.

68. The method of any one of claims 1-67, wherein the plurality of oligonucleotides is produced on a substrate.

69. The method of claim 68, wherein the plurality of oligonucleotides is produced in an array on a substrate.

70. The method of any one of claims 1-69, further comprising separating a subset of the DNA.

71. The method of any one of claims 1-70, wherein each oligonucleotide of a subset of the oligonucleotides comprises an index portion that is identical.

72. The method of any one of claims 1-71, wherein each transcribed DNA of a subset of the oligonucelotides comprises an index portion that is identical, the method further comprising separating the subset of the transcribed DNA based on the identical index portions.

73. A method, comprising:

simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides;

transcribing *in vitro* at least some of the amplified oligonucleotides to produce RNA;

reverse transcribing the RNA to produce transcribed DNA; and

selectively degrading the RNA relative to the transcribed DNA.

74. The method of claim 73, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides.

75. The method of any one of claims 73 or 74, wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences.

76. The method of any one of claims 73-75, wherein the plurality of oligonucleotides includes at least 100 unique oligonucleotide sequences.

77. The method of any one of claims 73-76, wherein the plurality of oligonucleotides includes at least 1,000 unique oligonucleotide sequences.

78. The method of any one of claims 73-77, wherein the plurality of oligonucleotides includes at least 10,000 unique oligonucleotide sequences.

79. The method of any one of claims 73-78, wherein the plurality of oligonucleotides includes at least 100,000 unique oligonucleotide sequences.

80. The method of any one of claims 73-79, wherein amplifying at least some of the plurality of oligonucleotides comprises exposing at least some of the plurality of oligonucleotides to primer-containing sequences.

81. The method of claim 80, wherein at least some of the primer-containing sequences further comprises a promoter, whereby the promoter is added to at least some of the amplified oligonucleotides.

82. The method of any one of claims 1-80, wherein at least some of the amplified oligonucleotides contain a promoter.

83. The method of any one of claims 81or 82, wherein the promoter is a T7 promoter.

84. The method of any one of claims 81 or 82, wherein the promoter is a T3 promoter.

85. The method of any one of claims 81or 82, wherein the promoter is a SP6 promoter.

86. The method of any one of claims 73-85, wherein at least some of the plurality of oligonucleotides comprise at least a first set of oligonucleotides having a first common index region, and a second set of oligonucleotides having a second common index region distinguishable from the first common index region.

87. The method of claim 86, comprising amplifying the first set of oligonucleotides but not the second set of oligonucleotides.

88. The method of any one of claims 86 or 87, wherein the plurality of oligonucleotides comprises at least 2 sets of oligonucleotides having distinguishable common index regions.

89. The method of any one of claims 86-88, wherein the plurality of oligonucleotides comprises at least 4 sets of oligonucleotides having distinguishable common index regions.

90. The method of any one of claims 86-89, wherein the plurality of oligonucleotides comprises at least 10 sets of oligonucleotides having distinguishable common index regions.

91. The method of any one of claims 86-90, wherein the plurality of oligonucleotides comprises at least 20 sets of oligonucleotides having distinguishable common index regions.

92. The method of any one of claims 86-91, wherein the plurality of oligonucleotides comprises at least 96 sets of oligonucleotides having distinguishable common index regions.

93. The method of any one of claims 86-92, wherein the plurality of oligonucleotides comprises at least 100 sets of oligonucleotides having distinguishable common index regions.

94. The method of any one of claims 86-93, wherein the plurality of oligonucleotides comprises at least 192 sets of oligonucleotides having distinguishable common index regions.

95. The method of any one of claims 73-94, wherein amplifying at least some of the plurality of oligonucleotides comprises producing amplified oligonucleotides comprising one oligonucleotide of the plurality of oligonucleotides and a promoter.

96. The method of any one of claims 73-95, wherein transcribing at least some of the amplified oligonucleotides comprises a mass of RNA that is at least 100-fold greater than the mass of amplified oligonucleotides.

97. The method of any one of claims 73-96, wherein transcribing at least some of the amplified oligonucleotides comprises exposing the amplified oligonucleotides to an RNA polymerase.

98. The method of claim 97, wherein the RNA polymerase is a T7 RNA polymerase.

99. The method of claim 97, wherein the RNA polymerase is a T3 RNA polymerase.

100. The method of claim 97, wherein the RNA polymerase is a SP6 RNA polymerase.

101. The method of claim 97, wherein the RNA polymerase is RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, or RNA polymerase V.

102. The method of claim 97, wherein the RNA polymerase is a viral RNA polymerase.

103. The method of any one of claims 73-102, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10 RNA copies of each of the amplified oligonucleotides.

104. The method of any one of claims 73-103, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 100 RNA copies of each of the amplified oligonucleotides.

105. The method of any one of claims 73-104, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 1,000 RNA copies of each of the amplified oligonucleotides.

106. The method of any one of claims 73-105, wherein transcribing at least some of amplified oligonucleotides to produce RNA comprises producing, on average, at least 10,000 RNA copies of each of the amplified oligonucleotides.

107. The method of any one of claims 73-106, wherein reverse transcribing the RNA comprises exposing the RNA to a reverse transcriptase.

108. The method of claim 107, wherein the reverse transcriptase is a viral reverse transcriptase.

109. The method of any one of claims 73-108, wherein reverse transcribing the RNA to produce transcribed DNA occurs without first purifying the RNA from components used to produce the RNA.

110. The method of any one of claims 73-109, further comprising purifying the RNA from components used to produce the RNA prior to reverse transcribing the RNA to produce transcribed DNA.

111. The method of any one of claims 73-110, wherein reverse transcribing the RNA to produce transcribed DNA comprises reverse transcribing the RNA to produce transcribed DNA using a sequence containing a transcription primer.

112. The method of claim 111, wherein the sequence containing a transcription primer is incorporated into the transcribed DNA.

113. The method of any one of claims 111 or 112, wherein the sequence containing a transcription primer further comprises a biotin moiety.

114. The method of any one of claims 111-113, wherein the sequence containing a transcription primer further comprises a digoxigenin moiety.

115. The method of any one of claims 111-114, wherein the sequence containing a transcription primer further comprises a fluorophore.

116. The method of any one of claims 111-115, wherein the sequence containing a transcription primer further comprises a non-nucleic acid moiety.

117. The method of any one of claims 111-116, wherein the sequence containing a transcription primer further comprises a signaling entity attached thereto.

118. The method of claim 117, wherein the signaling entity is fluorescent.

119. The method of any one of claims 117 or 118, wherein the signaling entity is photoswitchable between a first state able to emit light at a first wavelength, and a second state not able to emit light at the first wavelength.

120. The method of any one of claims 117-119, wherein the signaling entity is photactivatable.

121. The method of claim 120, wherein the signaling entity is a nucleic acid sequence.

122. The method of claim 121, wherein the additional nucleic acid sequence is able to hybridize selectively to a second nucleic acid sequence containing an additional signaling entity via Watson-Crick base-pairing.

123. The method of any one of claims 73-122, wherein selectively degrading the RNA relative to the transcribed DNA comprises chemically reducing the RNA.

124. The method of any one of claims 73-123, wherein selectively degrading the RNA relative to the transcribed DNA comprises exposing the RNA to a pH of at least about 9.

125. The method of any one of claims 73-124, wherein selectively degrading the RNA relative to the transcribed DNA comprises subjecting the RNA to alkaline hydrolysis.

126. The method of any one of claims 73-125, wherein selectively degrading the RNA relative to the transcribed DNA comprises enzymatically digesting the RNA.

127. The method of any one of claims 73-126, wherein the transcribed DNA is substantially single-stranded.

128. The method of any one of claims 73-127, wherein the steps are performed in the order recited.

129. The method of any one of claims 73-128, further comprising purifying the transcribed DNA.

130. The method of claim 129, comprising purifying the transcribed DNA by column purification.

131. The method of any one of claims 129 or 130, comprising purifying the transcribed DNA by ethanol precipitation.

132. The method of any one of claims 129-131, comprising purifying the transcribed DNA by solid-phase reversible immobilization.

133. The method of any one of claims 73-, further comprising concentrating the transcribed DNA.

134. The method of claim 133, comprising concentrating the transcribed DNA via evaporation.

135. The method of any one of claims 73-134, wherein the plurality of oligonucleotides is computationally designed.

136. The method of any one of claims 73-135, wherein the plurality of oligonucleotides is synthetically produced.

137. The method of any one of claims 73-136, wherein the plurality of oligonucleotides arises from a cell.

138. The method of any one of claims 73-137, wherein the plurality of oligonucleotides is produced on a substrate.

139. The method of claim 138, wherein the plurality of oligonucleotides is produced in an array on a substrate.

140. The method of any one of claims 73-139, further comprising separating a subset of the DNA.

141. The method of any one of claims 73-140, wherein each oligonucleotide of a subset of the oligonucleotides comprises an index portion that is identical.

142. The method of any one of claims 73-141, wherein each transcribed DNA of a subset of the oligonucelotides comprises an index portion that is identical, the method further comprising separating the subset of the transcribed DNA based on the identical index portions.

PTO/SB/17 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person are required to respond to a collection of information unless it displays a valid OMB control number

FEE TRANSMITTAL

Complete if known	
Application Number	Not Yet Assigned
Filing Date	Herewith
First Named Inventor	Xiaowei Zhuang
Examiner Name	N/A
Art Unit	N/A
Practitioner Docket No.	H0498.70510US00

[X] Applicant asserts small entity status. See 37 CFR 1.27
[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously

TOTAL AMOUNT OF PAYMENT ($) 130.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [X] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[X] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

[ ] Charge fee(s) indicated below
[ ] Charge fee(s) indicated below, **except for the filing fee**
[X] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17
[X] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	280	140*	70	600	300	150	720	360	180	
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	130.00

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims		**Extra Claims**		**Fee ($)**		**Fee Paid ($)**
_____	- or HP =	_____	x	_____	=	_____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
_____	_____

Indep. Claims		**Extra Claims**		**Fee ($)**		**Fee Paid ($)**
_____	- or HP =	_____	x	_____	=	_____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**		**Fee ($)**		**Fee Paid ($)**
66	- 100 =	_____	/50 =	_____	(round **up** to a whole number) x	_____	=	_____

4. OTHER FEE(S)

	Fees Paid ($)
Non-English specification, $130 fee (no small or micro entity discount)	
Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity)	
Other (e.g., late filing surcharge): _____	

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	September 15, 2014

3728772.1

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 62050636 Document Date: 09/15/2014

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

DocCode – APPENDIX

SCORE Placeholder Sheet for IFW Content

Application Number: 62050636 Document Date: 09/15/2014

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Appendix

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

DOCUMENT MADE AVAILABLE UNDER THE PATENT COOPERATION TREATY (PCT)

International application number:		**PCT/US2015/042556**
International filing date:		**29 July 2015 (29.07.2015)**
Document type:		**Certified copy of priority document**
Document details:	Country/Office:	**US**
	Number:	**62/050,636**
	Filing date:	**15 September 2014 (15.09.2014)**
Date of receipt at the International Bureau:		**03 August 2015 (03.08.2015)**

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a),(b) or (b-*bis*)

34, chemin des Colombettes
1211 Geneva 20, Switzerland
www.wipo.int

721260

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 02, 2015

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *62/142,653*
FILING DATE: *April 03, 2015*
RELATED PCT APPLICATION NUMBER: *PCT/US15/42556*

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS *US62/142,653*

Certified by

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 62142653 Document Date: 04/03/2015

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

PTO/SB/16 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

PROVISIONAL APPLICATION FOR PATENT COVER SHEET – Page 1 of 2

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. Via EFS-Web

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Xiaowei	Zhuang	Lexington, Massachusetts
Kok-Hao	Chen	Singapore, Republic of Singapore
Alistair	Boettiger	Cambridge, Massachusetts
Jeffrey R.	Moffitt	Somerville, Massachusetts
Siyuan	Wang	Cambridge, Massachusetts

Additional inventors are being named on the _____ separately numbered sheets attached hereto.

TITLE OF THE INVENTION (500 characters max):

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

[X] The address corresponding to Customer Number: 86110

OR

[ ] Firm or Individual Name

Address

City	State	Zip
Country	Telephone	Email

ENCLOSED APPLICATION PARTS (*check all that apply*)

[X] Application Data Sheet. See 37 CFR 1.76. [ ] CD(s), Number of CDs _____

[X] Drawing(s) *Number of Sheets* 17 [ ] Other (specify) _____

[X] Specification (e.g., description of the invention) *Number of Pages* 114

Fees Due: Filing Fee of $260 ($130 for small entity) ($65 for micro entity). If the specification and drawings exceed 100 sheets of paper, an application size fee is also due, which is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

METHOD OF PAYMENT OF THE FILING FEE AND APPLICATION SIZE FEE FOR THIS PROVISIONAL APPLICATION FOR PATENT

[X] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29.
Applicant must attach form PTO/SB/15A or B or equivalent.

[ ] A check or money order made payable to the *Director of the United States Patent and Trademark Office* is enclosed to cover the filing fee and application size fee (if applicable).

[X] Payment by credit card. Form PTO-2038 is attached.

130.00
TOTAL FEE AMOUNT ($)

[X] The Director is hereby authorized to charge the filing fee and application size fee (if applicable) or credit any overpayment to Deposit Account Number: 23/2825 .

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

3993437.1

PTO/SB/16 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

PROVISIONAL APPLICATION FOR PATENT COVER SHEET – Page 2 of 2

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☐ No.

☐ Yes, the invention was made by an agency of the U.S. Government. The U.S. Government agency name is: _____

☒ Yes, the invention was made under a contract with an agency of the U.S. Government. The name of the U.S. Government agency and Government contract number are:

National Institutes of Health - GM096450

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

SIGNATURE	/Tani Chen/	DATE	April 3, 2015
TYPED OR PRINTED NAME	Tani Chen, Sc.D.	REGISTRATION NO. *(if appropriate)*	52,728
TELEPHONE	617.646.8000	DOCKET NUMBER	H0498.70509US01

3993437.1

CLAIMS

1. A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2. The method of claim 1, comprising exposing the sample to at least 5 different nucleic acid probes.

3. The method of any one of claims 1 or 2, comprising exposing the sample to at least 10 different nucleic acid probes.

4. The method of any one of claims 1-3, comprising exposing the sample to at least 100 different nucleic acid probes.

5. The method of any one of claims 1-4, comprising exposing the sample to a plurality of nucleic acid probes simultaneously.

6. The method of any one of claims 1-5, comprising exposing the sample to a plurality of nucleic acid probes sequentially.

7. The method of claims 1-6, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8. The method of claim 7, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of RNA species in the sample.

9. The method of claim 7 or 8, wherein the combinatorial combination of nucleic acid probes target a combinatorial combination of DNA sequences in the sample.

10. The method of any one of claims 1-9, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11. The method of claim 10, wherein the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from the one or more read sequences.

12. The method of any one of claims 10 or 11, wherein the target sequence is substantially complementary to a nucleic acid sequence encoding a protein.

13. The method of any one of claims 10-12, wherein the target sequence is substantially complementary to messenger RNA (mRNA).

14. The method of any one of claims 10-13, wherein the plurality of nucleic acid probes comprises at least 8 possible read sequences.

15. The method of any one of claims 10-14, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

16. The method of any one of claims 10-15, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

17. The method of any one of claims 10-16, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

18. The method of any one of claims 10-17, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19. The method of any one of claims 10-18, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

20. The method of any one of claims 10-19, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

21. The method of any one of claims 10-20, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

22. The method of any one of claims 10-21, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

23. The method of any one of claims 10-22, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

24. The method of any one of claims 10-23, wherein at least some of the plurality of nucleic acid probes comprise no more than 3 read sequences.

25. The method of any one of claims 10-24, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

26. The method of claims 1-25, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27. The method of claim 26, wherein the first signaling entity is fluorescent.

28. The method of any one of claims 26 or 27, wherein the first signaling entity is a protein.

29. The method of any one of claims 26-28, wherein the first signaling entity is a dye.

30. The method of any one of claims 26-29, wherein the first signaling entity is a nanoparticle.

31. The method of any one of claims 26-30, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32. The method of claim 31, wherein the first signaling entity and the second signaling entity are identical.

33. The method of claim 31, wherein the first signaling entity and the second signaling entity are not identical.

34. The method of any one of claims 31-33, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35. The method of claim 34, comprising inactivating the first signaling entity by photobleaching at least some of the first signaling entity.

36. The method of any one of claims 34 or 35, comprising inactivating the first signaling entity by chemically bleaching at least some of the first signaling entity.

37. The method of any one of claims 34-36, comprising inactivating the first signaling entity by exposing the first signaling entity to a reactant able to alter the structure of the signaling entity.

38. The method of any one of claims 34-37, comprising inactivating the first signaling entity by removing at least some of the first signaling entity.

39. The method of any one of claims 34-38, comprising inactivating the first signaling entity by dissociating the first signaling entity from the first secondary probe.

40. The method of any one of claims 34-39, comprising inactivating the first signaling entity by dissociating the first secondary probe that contains the first signaling entity from the sample.

41. The method of any one of the claims 34-40, comprising inactivating the first signaling entity by chemically cleaving it from the first secondary probe.

42. The method of any one of the claims 34-41, comprising inactivating the first signaling entity by enzymatically cleaving it from the first secondary probe.

43. The method of any one of claims 34-42, comprising inactivating the first signaling entity by exposing the signaling entity or the first secondary probe to an enzyme.

44. The method of any one of claims 1-43, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

45. The method of any one of claims 1-44, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

46. The method of any one of claims 44 or 45, comprising determining the centroid using a maximum likelihood algorithm.

47. The method of any one of claims 44-46, comprising determining the centroid using a least squares algorithm.

48. The method of any one of claims 44-47, comprising determining the centroid using a Bayesian algorithm.

49. The method of any one of claims 44-48, comprising determining the centroid using a compressed sensing algorithm.

50. The method of any one of claims 1-49, further comprising determining a confidence level for the identified nucleic acid target.

51. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having one or more one-bit errors to the codeword.

52. The method of claim 51, comprising determining the confidence level using a ratio of the number of exact matches to the number of matches having exactly one one-bit error to the codeword.

53. The method of claims 1-52, wherein at least some of the plurality of nucleic acid probes comprise DNA.

54. The method of any one of claims 1-53, wherein at least some of the plurality of nucleic acid probes comprise RNA.

55. The method of any one of claims 1-54, wherein at least some of the plurality of nucleic acid probes comprise PNA.

56. The method of any one of claims 1-55, wherein at least some of the plurality of nucleic acid probes comprise LNA.

57. The method of any one of claims 1-56, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

58. The method of any one of claims 1-57, wherein at least some of the plurality of nucleic acids probes are configured to bind to a nucleic acid within the sample.

59. The method of any one of claims 1-58, wherein at least some of the binding of the nucleic acid probes to a target within the sample is specific binding.

60. The method of any one of claims 1-59, wherein at least some of the binding of the nucleic acid probes to a target within the sample is via Watson-Crick base pairing.

61. The method of any one of claims 1-60, wherein at least some of the plurality of nucleic acids probes are configured to bind to RNA.

62. The method of any one of claims 1-61, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

63. The method of any one of claims 1-62, wherein at least some of the plurality of nucleic acid probes are configured to bind to mRNA.

64. The method of any one of claims 1-63, wherein at least some of the plurality of nucleic acid probes are configured to bind to transfer RNA (tRNA).

65. The method of any one of claims 1-64, wherein at least some of the plurality of nucleic acid probes are configured to bind to ribosomal RNA (rRNA).

66. The method of any one of claims 1-65, wherein at least some of the plurality of nucleic acid probes are configured to bind to lncRNA.

67. The method of any one of claims 1-66, wherein at least some of the plurality of nucleic acid probes are configured to bind to snoRNA.

68. The method of any one of claims 1-67, wherein at least some of the plurality of nucleic acid probes are configured to bind to non-coding RNA.

69. The method of any one of claims 1-68, wherein the plurality of nucleic acids probes is configured to bind to DNA.

70. The method of any one of claims 1-5698, wherein at least some of the plurality of nucleic acid probes are configured to bind to genomic DNA.

71. The method of any one of claims 1-70, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

72. The method of any one of claims 1-71, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 100 nm.

73. The method of any one of claims 1-72, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 80 nm.

74. The method of any one of claims 1-73, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 50 nm.

75. The method of any one of claims 1-74, wherein the sample comprises a cell.

76. The method of claim 75, wherein the cell is a human cell.

77. The method of any one of claims 76 or 77, wherein the cell is fixed.

78. The method of any one of claims 1-77, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

79. The method of any one of claims 1-78, comprising determining binding of the nucleic acid probes using an optical imaging technique.

80. The method of any one of claims 1-79, comprising determining binding of the nucleic acid probes using a fluorescence imaging technique.

81. The method of any one of claims 1-80, comprising determining binding of the nucleic acid probes using a multi-color fluorescence imaging technique.

82. The method of any one of claims 1-81, comprising determining binding of the nucleic acid probes using a super-resolution fluorescence imaging technique.

83. The method of claim 82, comprising determining binding of the nucleic acid probes using stochastic optical reconstruction microscopy (STORM).

84. The method of any one of claims 82 or 83, comprising determining binding of the nucleic acid probes using photoactivated localization microscopy (PALM) or fluorescence photoactivation localization microscopy (FPALM).

85. The method of any one of claims 82-84, comprising determining binding of the nucleic acid probes using stimulated emission depletion microscopy (STED).

86. The method of any one of claims 82-85, comprising determining binding of the nucleic acid probes using structured illumination microscopy (SIM).

87. The method of any one of claims 82-7864, comprising determining binding of the nucleic acid probes using reversible saturated optically linear fluorescence transition (RESOLFT) microscopy.

88. A method, comprising:

exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and

for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

89. The method of claim 88, further comprising:

creating codewords based on determination of the read sequences within the sample.

90. The method of claim 89, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

91. A method, comprising:

exposing a sample to a plurality of primary nucleic acid probes;

exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary nucleic acid probes within the sample;

creating codewords based on fluorescence of the secondary nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

92. A method, comprising:

exposing a plurality of primary nucleic acid probes to a sample; and

exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample, wherein at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences.

93. The method of claim 92, further comprising:

creating codewords based on fluorescence of the secondary nucleic acid probes.

94. The method of claim 93, further comprising:

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

95. A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

96. The composition of claim 95, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

97. The composition of claim 96, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 3.

98. The composition of any one of claims 96 or 97, wherein the code is a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, or a Hamming(127, 120) code.

99. The composition of any one of claims 95-98, wherein the code is a SECDED code.

100. The composition of claim 99, wherein the code is a SECDED(8,4) code.

101. The composition of claim 99, wherein the code is a SECDED(16,4) code.

102. The composition of claim 99, wherein the code is a SECDED(16, 11) code, a SECDED (22, 16) code, a SECDED (39, 32) code, or a SECDED (72, 64) code.

103. The composition of any one of claims 99-90, wherein only codewords with a constant number of ones are used.

104. The composition of claim 99, wherein the code is MHD4 code.

105. The composition of claim 99, wherein the code is MHD2 code.

106. The composition of any one of claims 95-105, wherein the plurality of nucleic acid probes comprises no more than 100 possible read sequences.

107. The composition of any one of claims 95-106, wherein the plurality of nucleic acid probes comprises no more than 64 possible read sequences.

108. The composition of any one of claims 95-107, wherein the plurality of nucleic acid probes comprises no more than 32 possible read sequences.

109. The composition of any one of claims 95-108, wherein the plurality of nucleic acid probes comprises no more than 16 possible read sequences.

110. The composition of any one of claims 95-109, wherein the plurality of nucleic acid probes comprises no more than 8 possible read sequences.

111. The composition of any one of claims 95-110, wherein the target sequence of the plurality of nucleic acid probes has an average length of between 10 and 200 nucleotides.

112. The composition of any one of claims 95-111, wherein the plurality of read sequences have an average length of between 5 nucleotides and 50 nucleotides.

113. The composition of any one of claims 95-112, wherein at least some of the plurality of nucleic acid probes comprise no more than 10 read sequences.

114. The composition of any one of claims 95-113, wherein at least some of the plurality of nucleic acid probes comprise no more than 6 read sequences.

115. The composition of any one of claims 95-114, wherein at least some of the plurality of nucleic acid probes comprise no more than 4 read sequences.

116. The composition of any one of claims 95-115, wherein at least some of the plurality of nucleic acid probes comprise no more than 2 read sequences.

117. The composition of any one of claims 95-116, wherein at least some of the plurality of nucleic acid probes comprise DNA.

118. The composition of any one of claims 95-117, wherein at least some of the plurality of nucleic acid probes comprise RNA.

119. The composition of any one of claims 95-118, wherein at least some of the plurality of nucleic acid probes comprise PNA.

120. The composition of any one of claims 95-119, wherein at least some of the plurality of nucleic acid probes comprise LNA.

121. The composition of any one of claims 95-120, wherein the plurality of nucleic acid probes have an average length of between 10 and 300 nucleotides.

122. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

123. The method of claim 122, wherein the code space has a Hamming distance of at least 2.

124. The method of any one of claims 122 or 123, wherein the plurality of targets comprise RNA.

125. A method, comprising:

exposing nucleic acid probes formed using the method of any one of claims 108-110 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets within

the sample.

126. The method of claim 125, wherein at least some of the plurality of secondary nucleic probes is able to bind to a read sequence of a nucleic acid probe.

127. The method of claim 126, wherein at least some of the plurality of secondary nucleic probes is able to specifically bind to only one read sequence.

128. The method of claim 125-127, wherein if no match is found between a valid codeword and the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

129. The method of claim 125-128, wherein at least some of the secondary nucleic acid probes each comprise a signaling entity.

130. The method of claim 129, comprising imaging the signaling entity to determine binding of the plurality of secondary nucleic probes within the sample.

131. The method of claim 130, comprising imaging the signaling entity using a super-resolution imaging technique.

132. The method of any one of claims 130 or 131, further comprising applying drift correction to the images.

133. The method of any one of claims 130-132, further comprising applying physical registration to the images.

134. The method of claim 125-133, comprising determining binding of the plurality of secondary nucleic probes within the sample at a resolution better than 100 nm.

135. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

136. The method of claim 135, where in the codewords comprise a number of positions that is less than the number of targets.

137. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

138. The method of claim 137, wherein the codewords form an error-checking space.

139. The method of any one of claims 137 or 138, wherein the codewords form an error-correcting code space.

140. A method, comprising:

associating a plurality of targets with a plurality of target sequences and a

plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes each comprising a target sequence; and

forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

141. The method of any one of claims 137-140, wherein the groups are applied to the sample sequentially using fluidic devices.

142. The method of any one of claims 137-141, wherein each group of nucleic acid probes contains a group identification sequence distinguishable from the group identification sequence of the other groups.

143. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code;

associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample;

forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences;

forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences;

exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample;

exposing the primary nucleic acid probes in the sample to a secondary nucleic

acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

144. The method of claim 143, wherein the wherein the read sequences are taken from a set of orthogonal sequences, which have a homology of less than 15 basepairs with one another and with the nucleic acid species in a sample.

145. The method of any one of claims 143 or 144, comprising combining the images of the sample into a combined image.

146. The method of claim 145, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

147. The method of any one of claims 143-146, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

148. The method of any one of claims 143-147, comprising determining the centroid of the first signaling entity using an algorithm for determining partially overlapping single emitters.

149. The method of any one of claims 143-148, comprising determining the centroid using a maximum likelihood algorithm.

150. The method of any one of claims 143-149, comprising determining the centroid using a least squares algorithm.

151. The method of any one of claims 143-150, comprising determining the centroid using a Bayesian algorithm.

152 The method of any one of claims 143-151, comprising determining the centroid using a compressed sensing algorithm.

153. The method of any one of claims 146-152, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

154. The method of claim 153, comprising matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword that correspond to a nucleic acid target

155. The method of any one of claims 143-154, comprising imaging the sample using fluorescence microscopy.

156. The method of any one of claims 143-155, comprising imaging the sample at a resolution better than 500 nm.

157. The method of any one of claims 143-156, comprising imaging the sample at a resolution better than 300 nm.

158. The method of any one of claims 143-157, comprising imaging the sample at a resolution better than 100 nm.

159. The method of any one of claims 143-158, comprising imaging the sample at a resolution better than 80 nm.

160. The method of any one of claims 143-159, comprising imaging the sample at a resolution better than 50 nm.

161. The method of any one of claims 143-160, comprising imaging the sample at a resolution better than 20 nm.

162. The method of any one of claims 143-161, comprising imaging the sample using a super-resolution fluorescence imaging technique.

163. The method of any one of claims 143-162, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

164. The method of any one of claims 143-163, where in the codewords comprise a number of positions that is less than the number of targets.

165. A method, comprising:

associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets;

grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords;

exposing a sample to one of the probe pools;

imaging the sample; and

repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

166. The method of claim 165, comprising combining the images of the sample into a combined image.

167. The method of claim 166, further comprising identifying foreground features corresponding to putative nucleic acid targets in the sample.

168. The method of claim 167, comprising determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample.

169. The method of claim 168, comprising matching the determined codeword to a valid codeword wherein, if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target.

170. The method of any one of claims 165-169, comprising imaging the sample using fluorescence microscopy.

171. The method of any one of claims 165-170, comprising imaging the sample at a resolution better than 300 nm.

172. The method of any one of claims 165-171, comprising imaging the sample at a resolution better than 100 nm.

173. The method of any one of claims 165-172, comprising imaging the sample using a super-resolution fluorescence imaging technique.

174. The method of any one of claims 165-173, comprising imaging the sample using a technique selected from the group consisting of STORM, PALM, FPALM, STED, SIM, RESOLFT, SOFI or SPDM.

175. The method of any one of claims 165-174, wherein the codewords comprise a number of positions that is less than the number of targets.

Electronic Acknowledgement Receipt	
EFS ID:	21966075
Application Number:	62142653
International Application Number:	
Confirmation Number:	5966
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Angela Griffith
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US01
Receipt Date:	03-APR-2015
Filing Date:	
Time Stamp:	14:06:42
Application Type:	Provisional

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$130
RAM confirmation Number	325
Deposit Account	232825
Authorized User	CHEN, TANI
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees) Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Provisional Cover Sheet (SB16)	H049870509US01-PROCVR-TC.pdf	38743 453e9106fcfe3eafa433f9b83e908df6e624e390	no	2
Warnings:					
This is not a USPTO supplied Provisional Cover Sheet SB16 form.					
Information:					
2	Fee Worksheet (SB06)	H049870509US01-FEE-TC.pdf	29936 dc476dcfa2751d00e733494e3615a6ea7c8b6b53	no	1
Warnings:					
Information:					
3	Application Data Sheet	H049870509US01-ADS-TC.pdf	23135 8cc62ae3bde1143e167008d33e3a12e7fa4585a8	no	6
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
4		H049870509US01-PROAPL-TC.pdf	394026 e8511b1e4b9661be031c4780d25d87aea0f6b41a	yes	114

Multipart Description/PDF files in .zip description

Document Description	Start	End
Specification	1	92
Claims	93	113
Abstract	114	114

Warnings:					
Information:					
5	Drawings-other than black and white line drawings	H049870509US01-DRW-TC.pdf	2544673 bbc1402220047b6059b180c18d9920e522767619	no	17
Warnings:					
Information:					

6	Fee Worksheet (SB06)	fee-info.pdf	30270 0fa78ae890c64c4fbfa11beb6f8893717c5c873e	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	3060783	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Application Data Sheet

Inventor Information

Inventor Number::	1
Given Name::	Xiaowei
Family Name::	Zhuang
City of Residence::	Lexington
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	5 Luongo Farm Lane
City of mailing address::	Lexington
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02421

Inventor Number::	2
Given Name::	Kok-Hao
Family Name::	Chen
City of Residence::	Singapore
Country of Residence::	Republic of Singapore
Street of mailing address::	37 Jalan Lepas
City of mailing address::	Singapore
Country of mailing address::	Republic of Singapore
Postal or Zip Code of mailing address::	537961

Inventor Number::	3
Given Name::	Alistair
Family Name::	Boettiger
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	1550 Cambridge Street
	Apt. #2
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02139

Inventor Number::	4
Given Name::	Jeffrey
Middle Name::	R.
Family Name::	Moffitt
City of Residence::	Somerville
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	175 Beacon Street
	Apt. #108
City of mailing address::	Somerville
State or Province of mailing address::	MA

Postal or Zip Code of mailing address::	02143
Inventor Number::	5
Given Name::	Siyuan
Family Name::	Wang
City of Residence::	Cambridge
State or Province of Residence::	MA
Country of Residence::	US
Street of mailing address::	28 Fernald Drive
	Apt. #31
City of mailing address::	Cambridge
State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02138

Correspondence Information

Correspondence Customer Number::	86110

Application Information

Application Type::	Provisional
Subject Matter::	Utility
CD-ROM or CD-R?::	None
Sequence submission?::	None
Computer Readable Form (CRF)?::	No

Title::	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Attorney Docket Number::	H0498.70509US01
Request for Early Publication?::	No
Request for Non-Publication?::	No
Small Entity?::	Yes
Petition included?::	No
Licensed US Govt. Agency::	National Institutes of Health
Contract or Grant Numbers::	GM096450

Representative Information

Representative Customer Number::	86110

Domestic Priority Information

Foreign Priority Information

Applicant Information

Applicant Type::	Person to whom inventors are obligated to assign
Organization Name::	President and Fellows of Harvard College
Street of mailing address::	17 Quincy Street
City of mailing address::	Cambridge

State or Province of mailing address::	MA
Postal or Zip Code of mailing address::	02138

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.			
Signature	/Tani Chen/	Date	April 3, 2015
Name (Print/Type)	Tani Chen, Sc.D.	Registration No. (Attorney/Agent)	52,728

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a

second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each

codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution

better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Fig. 12 shows schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16B show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention;

Figs. 19A-9D illustrate decoding and error assessment, in another embodiment of the invention; and

Fig. 20 illustrates various examples of codebooks, in other embodiments of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGATC

ATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCGGACA
AACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;
SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;
SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;
SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;
SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;
SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;
SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;
SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and
SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may

contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences. For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one

secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second

secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling

entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or

different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein. Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human

cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or

simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less

than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer

sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the

determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary

nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments, the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences

described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form

a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with

"0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about 500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about 0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about

600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7,

etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a

suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700

(Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission

portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate, Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate,

Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF,

Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after any human has finished taking any action to promote the function, e.g. by entering instructions into a computer to start the process. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

For example, in some cases, a computer may be used to control imaging of the sample, e.g., using fluorescence microscopy, STORM or other super-resolution techniques such as those described herein. In some cases, the computer may also control operations such as drift correction, physical registration, hybridization and cluster alignment in image analysis, cluster decoding (e.g., fluorescent cluster decoding), error detection or correction (e.g., as discussed herein), noise reduction, identification of foreground features from background features (such as noise or debris in images), or the like. As an example, the computer may be used to control activation and/or excitation of signaling entities within the sample, and/or the acquisition of images of the signaling entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the signaling entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of signaling entities in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may be used to construct an image and/or determine the locations of the signaling entities, in some cases at high resolutions, as noted above.

In some aspects, the sample is positioned on a microscope. In some cases, the microscope may contain one or more channels, such as microfluidic channels, to direct or control fluid to or from the sample. For instance, in one embodiment, nucleic acid probes such as those discussed herein may be introduced and/or removed from the sample by flowing fluid through one or more channels to or from the sample. In some cases, there may also be one or more chambers or reservoirs for holding fluid, e.g., in fluidic communication with the channel, and/or

with the sample. Those of ordinary skill in the art will be familiar with channels, including microfluidic channels, for moving fluid to or from a sample.

As used herein, "microfluidic," "microscopic," "microscale," the "micro-" prefix (for example, as in "microchannel"), and the like generally refers to elements or articles having widths or diameters of less than about 1 mm, and less than about 100 microns (micrometers) in some cases. In some embodiments, larger channels may be used instead of, or in conjunction with, microfluidic channels for any of the embodiments discussed herein. For examples, channels having widths or diameters of less than about 10 mm, less than about 9 mm, less than about 8 mm, less than about 7 mm, less than about 6 mm, less than about 5 mm, less than about 4 mm, less than about 3 mm, or less than about 2 mm may be used in certain instances. In some cases, the element or article includes a channel through which a fluid can flow. In all embodiments, specified widths can be a smallest width (i.e. a width as specified where, at that location, the article can have a larger width in a different dimension), or a largest width (i.e. where, at that location, the article has a width that is no wider than as specified, but can have a length that is greater). Thus, for instance, the microfluidic channel may have an average cross-sectional dimension (e.g., perpendicular to the direction of flow of fluid in the microfluidic channel) of less than about 1 mm, less than about 500 microns, less than about 300 microns, or less than about 100 microns. In some cases, the microfluidic channel may have an average diameter of less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 5 microns, less than about 3 microns, or less than about 1 micron.

A "channel," as used herein, means a feature on or in an article (e.g., a substrate) that at least partially directs the flow of a fluid. In some cases, the channel may be formed, at least in part, by a single component, e.g. an etched substrate or molded unit. The channel can have any cross-sectional shape, for example, circular, oval, triangular, irregular, square or rectangular (having any aspect ratio), or the like, and can be covered or uncovered (i.e., open to the external environment surrounding the channel). In embodiments where the channel is completely covered, at least one portion of the channel can have a cross-section that is completely enclosed, and/or the entire channel may be completely enclosed along its entire length with the exception of its inlet and outlet.

A channel may have any aspect ratio, e.g., an aspect ratio (length to average cross-sectional dimension) of at least about 2:1, more typically at least about 3:1, at least about 5:1, at least about 10:1, etc. As used herein, a "cross-sectional dimension," in reference to a fluidic or microfluidic channel, is measured in a direction generally perpendicular to fluid flow within the channel. A channel generally will include characteristics that facilitate control over fluid transport, e.g., structural characteristics and/or physical or chemical characteristics (hydrophobicity vs. hydrophilicity) and/or other characteristics that can exert a force (e.g., a containing force) on a fluid. The fluid within the channel may partially or completely fill the channel. In some cases the fluid may be held or confined within the channel or a portion of the channel in some fashion, for example, using surface tension (e.g., such that the fluid is held within the channel within a meniscus, such as a concave or convex meniscus). In an article or substrate, some (or all) of the channels may be of a particular size or less, for example, having a largest dimension perpendicular to fluid flow of less than about 5 mm, less than about 2 mm, less than about 1 mm, less than about 500 microns, less than about 200 microns, less than about 100 microns, less than about 60 microns, less than about 50 microns, less than about 40 microns, less than about 30 microns, less than about 25 microns, less than about 10 microns, less than about 3 microns, less than about 1 micron, less than about 300 nm, less than about 100 nm, less than about 30 nm, or less than about 10 nm or less in some cases. In one embodiment, the channel is a capillary.

A variety of materials and methods, according to certain aspects of the invention, can be used to form devices or components containing microfluidic channels, chambers, etc. For example, various devices or components can be formed from solid materials, in which the channels can be formed via micromachining, film deposition processes such as spin coating and chemical vapor deposition, physical vapor deposition, laser fabrication, photolithographic techniques, etching methods including wet chemical or plasma processes, electrodeposition, and the like. See, for example, *Scientific American*, 248:44-55, 1983 (Angell, *et al*).

In one set of embodiments, various structures or components can be formed of a polymer, for example, an elastomeric polymer such as polydimethylsiloxane ("PDMS"), polytetrafluoroethylene ("PTFE" or Teflon®), or the like. For instance, according to one embodiment, a channel such as a microfluidic channel may be implemented by fabricating the fluidic system separately using PDMS or other soft lithography techniques (details of soft

lithography techniques suitable for this embodiment are discussed in the references entitled "Soft Lithography," by Younan Xia and George M. Whitesides, published in the *Annual Review of Material Science*, 1998, Vol. 28, pages 153-184, and "Soft Lithography in Biology and Biochemistry," by George M. Whitesides, Emanuele Ostuni, Shuichi Takayama, Xingyu Jiang and Donald E. Ingber, published in the *Annual Review of Biomedical Engineering*, 2001, Vol. 3, pages 335-373; each of these references is incorporated herein by reference).

Other examples of potentially suitable polymers include, but are not limited to, polyethylene terephthalate (PET), polyacrylate, polymethacrylate, polycarbonate, polystyrene, polyethylene, polypropylene, polyvinylchloride, cyclic olefin copolymer (COC), polytetrafluoroethylene, a fluorinated polymer, a silicone such as polydimethylsiloxane, polyvinylidene chloride, bis-benzocyclobutene ("BCB"), a polyimide, a fluorinated derivative of a polyimide, or the like. Combinations, copolymers, or blends involving polymers including those described above are also envisioned. The device may also be formed from composite materials, for example, a composite of a polymer and a semiconductor material.

In some embodiments, various microfluidic structures or components of the device are fabricated from polymeric and/or flexible and/or elastomeric materials, and can be conveniently formed of a hardenable fluid, facilitating fabrication via molding (e.g. replica molding, injection molding, cast molding, etc.). The hardenable fluid can be essentially any fluid that can be induced to solidify, or that spontaneously solidifies, into a solid capable of containing and/or transporting fluids contemplated for use in and with the fluidic network. In one embodiment, the hardenable fluid comprises a polymeric liquid or a liquid polymeric precursor (i.e. a "prepolymer"). Suitable polymeric liquids can include, for example, thermoplastic polymers, thermoset polymers, waxes, metals, or mixtures or composites thereof heated above their melting point. As another example, a suitable polymeric liquid may include a solution of one or more polymers in a suitable solvent, which solution forms a solid polymeric material upon removal of the solvent, for example, by evaporation. Such polymeric materials, which can be solidified from, for example, a melt state or by solvent evaporation, are well known to those of ordinary skill in the art. A variety of polymeric materials, many of which are elastomeric, are suitable, and are also suitable for forming molds or mold masters, for embodiments where one or both of the mold masters is composed of an elastomeric material. A non-limiting list of examples of such polymers includes polymers of the general classes of silicone polymers, epoxy polymers,

and acrylate polymers. Epoxy polymers are characterized by the presence of a three-membered cyclic ether group commonly referred to as an epoxy group, 1,2-epoxide, or oxirane. For example, diglycidyl ethers of bisphenol A can be used, in addition to compounds based on aromatic amine, triazine, and cycloaliphatic backbones. Another example includes the well-known Novolac polymers. Non-limiting examples of silicone elastomers suitable for use according to the invention include those formed from precursors including the chlorosilanes such as methylchlorosilanes, ethylchlorosilanes, phenylchlorosilanes, etc.

Silicone polymers are used in certain embodiments, for example, the silicone elastomer polydimethylsiloxane. Non-limiting examples of PDMS polymers include those sold under the trademark Sylgard by Dow Chemical Co., Midland, MI, and particularly Sylgard 182, Sylgard 184, and Sylgard 186. Silicone polymers including PDMS have several beneficial properties simplifying fabrication of various structures of the invention. For instance, such materials are inexpensive, readily available, and can be solidified from a prepolymeric liquid via curing with heat. For example, PDMSs are typically curable by exposure of the prepolymeric liquid to temperatures of about, for example, about 65 °C to about 75 °C for exposure times of, for example, at least about an hour. Also, silicone polymers, such as PDMS, can be elastomeric and thus may be useful for forming very small features with relatively high aspect ratios, necessary in certain embodiments of the invention. Flexible (e.g., elastomeric) molds or masters can be advantageous in this regard.

One advantage of forming structures such as microfluidic structures or channels from silicone polymers, such as PDMS, is the ability of such polymers to be oxidized, for example by exposure to an oxygen-containing plasma such as an air plasma, so that the oxidized structures contain, at their surface, chemical groups capable of cross-linking to other oxidized silicone polymer surfaces or to the oxidized surfaces of a variety of other polymeric and non-polymeric materials. Thus, structures can be fabricated and then oxidized and essentially irreversibly sealed to other silicone polymer surfaces, or to the surfaces of other substrates reactive with the oxidized silicone polymer surfaces, without the need for separate adhesives or other sealing means. In most cases, sealing can be completed simply by contacting an oxidized silicone surface to another surface without the need to apply auxiliary pressure to form the seal. That is, the pre-oxidized silicone surface acts as a contact adhesive against suitable mating surfaces. Specifically, in addition to being irreversibly sealable to itself, oxidized silicone such as oxidized

PDMS can also be sealed irreversibly to a range of oxidized materials other than itself including, for example, glass, silicon, silicon oxide, quartz, silicon nitride, polyethylene, polystyrene, glassy carbon, and epoxy polymers, which have been oxidized in a similar fashion to the PDMS surface (for example, via exposure to an oxygen-containing plasma). Oxidation and sealing methods useful in the context of the present invention, as well as overall molding techniques, are described in the art, for example, in an article entitled "Rapid Prototyping of Microfluidic Systems and Polydimethylsiloxane," *Anal. Chem.*, 70:474-480, 1998 (Duffy *et al.*), incorporated herein by reference.

The following documents are each incorporated herein by reference in their entireties: U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; and Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

The example presents a platform to enable the simultaneous detection of the number and spatial organization of thousands of distinct mRNAs within single cells with high efficiency and low error-rate using a novel form of highly multiplexed fluorescence *in situ* hybridization (FISH). This example accomplishes these measurements by integrating and innovating methods for massively parallel probe synthesis, super-resolution imaging, and self-correcting error-checking codes.

Here, these examples present methods for the simultaneous detection of some or all of the thousands of unique RNAs expressed in a cell. This approach not only promises to revolutionize the throughput of the already effective single-molecule FISH (smFISH) approach, but also allows researchers to benefit from the hypothesis free discovery approach which has made other whole-genome systems approaches to biology so effective. For example, this whole genome approach may allow researchers to discover RNAs whose expression levels and/or subcellular localization patterns change under certain conditions of interest, such as disease states, without knowing, *a priori*, which mRNA will change in abundance or localization. Simultaneous

measurements of hundreds of genes within a single cell also allow for the identification of correlations between genes in expression and localization patterns in some cases.

This can be achieved using methods for highly multiplexed smFISH via the sequential hybridization of orthogonal detection probes and super-resolution imaging, reducing the cost of probe synthesis, and the development of a highly automated system to minimize demands on the user, as discussed herein. This provides an integrated platform to handle the bioinformatics of probe design, the mathematics of error-correcting codes, the complexity of image registration and analysis, and the cumbersome fluid handling through a simple suite of user friendly interfaces. This integration allows easy operation with limited user training and facilitates the rapid collection of data.

This example illustrates: (1) computational design of "codewords" to attach to all RNA targets in the cell that will allow unique identification of each RNA with some degree of experimental error tolerance, (2) translation of these codewords into nucleotide sequences and synthesis of the required single-stranded (ss) oligonucleotide (e.g. ssDNA) probes, (3) sample fixation and hybridization of these probes to the RNA targets *in situ*, (4) read-out of these codewords via successive rounds of hybridization of distinct fluorescent probes imaged with conventional fluorescence microscopy or super-resolution fluorescence microscopy, and (5) automated decoding of measured codewords combined with computational error correction to uniquely and robustly identify individual mRNAs.

In the first step, a "codeword" is assigned to every RNA that is to be labeled. In a typical design these may be strings of N binary letters or positions. Codewords may be chosen from the same wide range of existing error tolerant or error-correcting encoding schemes developed for digital storage and communication, e.g., using Hamming codes or the like. For example, actin-RNA may be assigned the binary codeword 11001010. Each codeword may be unique and separated from the other codewords by a Hamming distance, h, which measures the number of letters or positions that must be incorrectly read for one codeword to be misinterpreted as a different one. A Hamming distance greater than 1 between all codewords allows for some measurement errors to be detected—since simple errors would produce codewords that are not used to encode RNAs. For a Hamming distance larger than 2, it is also possible to correct some errors, as codewords with one error will be closest in Hamming distance to a single, unique codeword. The total number of different RNAs to be detected from the transcriptome and the

amount of error correction desired determines the length of the codewords. Information theory provides several efficient algorithms for assembling error-correcting binary codebooks.

In the second step, this encoding scheme is translated into a set of oligonucleotide (e.g. DNA) probe sequences, which can be called primary probes or encoding probes, each of which not only targets a probe to the RNA of interest but also encodes the unique binary codeword within a set of secondary binding sites (Fig. 1). For example, first designed may be primary binding sequences for each targeted mRNA. These sequences are "target sequences" that are comprised of complementary nucleotide sequences to their target RNAs computationally selected to satisfy a stringent set of hybridization conditions, including uniqueness in the target genome. To improve the efficiency of hybridization to individual mRNAs, multiple primary target sequences are designed for each individual RNA. Then each position within the set of codewords is assigned a unique oligonucleotide (e.g. DNA) sequence, which is called a read sequence. These tags are designed as to have no interaction with endogenous mRNA sequences or each other. For instance, for all the value "1"s in a codeword of an individual mRNA, the corresponding read sequence is attached to the primary targeting sequences against that mRNA. In general, each probe will contain a target sequence and one or more read sequences. If the total length of the necessary read sequences and the primary target sequence exceeds synthesis capabilities, then subsets of the read sequences can be appended to distinct target sequences. For example, consider the potential codeword 11001010 for actin. Probe sequences for this RNA could contain the read sequences corresponding to positions 1, 2, 5, and 7 in the codeword attached to a variety of target sequences specific to actin. After all the sequences have been designed, the resulting complex set of unique custom oligonucleotide (e.g. DNA) sequences is manufactured and amplified using methods as described below.

In the third step, the resulting pool of DNA is hybridized, e.g., to fixed, permeabilized cells. In this process, individual probes may be attached to every RNA in the cell by hybridization of their corresponding target sequences with the RNA while the read sequences remain free to bind the appropriate secondary probes as discussed below.

In the fourth step—the read-out step—fluorescently labeled secondary nucleic acid probes (also called readout probes) are successively hybridized to the read sequences attached to the target sequences that binds to the mRNA targets in the above step. When simultaneously imaging a large number of different RNA species in cells, the density of labeled RNAs may

exceed that at which each RNA can be resolved via conventional imaging methods. Thus, this may be performed using a super-resolution imaging method, for example STORM (stochastic optical reconstruction microscopy), to resolve the labeled molecules. After each round of hybridization and imaging with the secondary probes, the fluorophores are quenched or otherwise inactivated either via chemical or optical techniques such as oxidation, chemically bleaching, photobleaching, stringent washing or enzymatic digestion, etc. The sample is then stained with the next secondary probe, and the cycle continues until all positions of the codewords have been read out. In the simplest incarnation, there will be one hybridization step for each position within the codeword, e.g. 8 hybridization steps for an 8-letter codeword (Fig. 1).

Fig. 1 shows schematic diagrams of this example. Fig. 1A shows that every position of the codewords is assigned a unique oligonucleotide sequence when this position has a value "1." All mRNA codewords are then translated into combinations of read sequences, which are attached to the targeting sequence. Fig. 1B shows various steps of the labeling scheme of this example. In the first step, all mRNAs (I - III) are tagged with multiple oligonucleotide (e.g. ssDNA) probes comprising a primary targeting sequence, which hybridizes to the RNA of interest, and a "tail" (i.e. containing read sequences) carrying the translated codeword, which does not interact with endogenous nucleotide sequences. In the next step, the first secondary probe is added, which can bind all probes whose tails have a read sequence corresponding to the value of "1" in the first position. The dyes on these secondary probes are imaged and bleached, then the next secondary probe is added to bind probes attached to mRNA which have a value of "1" in the second position of their assigned codeword, and so on.

In the final step, the microscopy images from each staining and imaging round are aligned, for example, computationally (e.g. using fiducial beads or other markers tracked during image acquisition), and the clusters of localizations resolved by conventional fluorescence microscopy or super-resolution imaging (e.g. STORM) from the different rounds are identified. These clusters of localizations arise from individual target mRNA molecules, and the hybridization rounds in which a spot was detected in a given cluster correspond to the "1" in the codeword for that mRNA. If there are no missed-detection events or false positive signal in the images, this codeword will perfectly match one of the expected codewords. Fig. 1 describes an example in which the codeword has three letters, i.e. three positions, and the three target mRNAs

have codewords 110, 101, and 011 assigned to them. In real experimental examples, the codeword could contain more digits. For example, the mRNA for actin can be assigned the codeword, 11001010. In that case, detected clusters containing overlapping localization signals in the 1st, 2nd, 5th and 7th hybridization steps (meaning the 1st, 2nd, 5th and 7th secondary probes bound to this site) can be identified as individual actin mRNA molecules, since the pattern of positive bindings matches the codewode of actin (11001010). In addition, if there are missed-detection events or false positive signals in the image data, these aberrations can be corrected by the implemented error-correction scheme. For example, clusters of localizations with a detected codeword that has only one digit discrepancy from 11001010 (such as 11000010 or 11101010) can also be identified as actin mRNA since all other valid codeword in this example differ from the detected pattern in two more positions.

EXAMPLE 2

This example describes another alternative approach that differs in several of the steps described above. This approach begins with the first step, construction of the codewords to the desired mRNA targets, as described above.

In the second step of this approach, nucleic acid probes are designed that bind uniquely to the mRNA targets of interest, as described above. However, instead of appending unique read sequences to these targeting sequences, unique pools or groups of probes are constructed from these target sequences. Each pool comprises all or a subset of the sequences that target all mRNAs which contain the same value at a given position in their codeword. For example, the first pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the first position of their codewords, e.g. 110 and 101 but not 011; the second pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the second position of their codewords, e.g. 110 and 011 but not 101; the third pool would have of all or a subset of the target sequences designed for all mRNAs that contain a 1 in the third position of their codewords, e.g. 011 and 101 but not 110 (Fig. 1C). As another example, consider the potential codeword 11001010 for actin. Probes that target this mRNA would be included in pools 1, 2, 5, and 7 but not in pools 3, 4, 6, and 8. The same target to a given mRNA may or may not be included in pools. For example, a probe that targets the same region of actin may be included in pools 1, 2, 5, and 7 or any subset of these pools. After all of the pools have

been designed, each complex set of unique, custom oligonucleotide sequences is manufactured and amplified using methods as described below.

In the third step of this approach, the first pool of probes is hybridized, e.g., to fixed, permeabilized cells. In this process, the fluorophores attached to each of the probes in this pool are bound to each of the targets of that pool. The binding of these probes is then determined by fluorescence microscopy. As described above, these images can be collected either via a range of methods including both conventional fluorescence imaging or super-resolution imaging methods such as STORM. After a round of imaging, the probes from the first pool are inactivated or removed from the sample via the methods described above. This process is then repeated for each successive pool of probes until some or all of the pools have been applied to the sample and imaged such that all positions in the codewords have been read out. In the simplest incarnation, there will be one hybridization and imaging step for each position in the codeword, e.g. 3 rounds of hybridization and imaging for a codeword with 3 positions (Fig 1C) or 8 rounds of hybridization and imaging for a codeword with 8 positions.

The final step of this approach is identical to that described above.

EXAMPLE 3

In this example, 14 genes (PGK1, H3F3B, PKM, ENO1, GPI, EEF2, GNAS, HSPA8, GAPDH, CALM1, RHOA, PPIA, UBA52, and VCP) were encoded using a subset of the (8,4) SECDED code (Figs. 2A-2E). To determine the accuracy of these measurements, the measured abundances of these 14 mRNAs were compared to the abundances measured from bulk RNA-seq of A549 cells (published ENCODE data). Remarkably, it was found that there was excellent agreement between these two measurements, as the transcripts count measured using the sequential hybridization approach correlated with gene expression measured using RNA-seq with a Pearson correlation coefficient r of 0.75 (Fig. 2F). Gene expression from 3 other cells were also measured, and it was found that the gene expression of these 14 genes was highly correlated among the cells with a r of 0.96 (Fig. 2G).

Codebook Design. Each mRNA in the target set was assigned a binary codeword using a Single Error Correction Double Error Detection (SECDED) code. SECDED is an extended Hamming codebook with an additional parity bit. Briefly, Matlab's Communications System toolbox was used to generate SECDED codes of either 8 or 16 letters or positions. In both cases, only those codewords containing four 1s were used. These words were assigned at random to

mRNAs in the target set. [0 1 0 1 1 1 0 0] is an example of the 8-letter codewords used (i.e., these codewords each contained four 1s and four 0s.) [0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0] is an example of the 16-letter codewords used (i.e., each codeword contained four 1s and twelve 0s). Not every codeword was necessarily assigned to an mRNA.

Computational Assembly of ssDNA Primary Probe Sequence. The number of primary nucleic acid probes used for hybridization with mRNA targets ranged from 200 to 2000 unique oligonucelotides, depending on the experiment. For example, to label 14 mRNAs with 28 oligos targeting each gene, 392 unique sequences were used. Large number of oligos with unique sequences were purchased in a pool from LC Sciences or CustomArray. However, array synthesized oligos were in minute quantity that was insufficient for *in situ* hybridization. The protocol for their amplification are described below.

Each primary probe contained three components: flanking primer sequences to allow enzymatic amplification of probes, targeting sequence for *in situ* hybridization to mRNAs, and secondary tag sequence containing one or more read sequences for sequential readout of codewords.

The following is an example of a primary probe:

GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCG
CTTAATGGATCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTG
CGTCGGTCGGACAAACGAGG

(SEQ ID NO: 1)

The components are arranged in the following order: forward primer (not underlined), secondary read sequence 1 (underlined), mRNA targeting sequence (not underlined), secondary read sequence 2 (underlined), and reverse primer (not underlined). The secondary read sequences are the reverse complement of the corresponding secondary probes. Since only codewords that contained four '1's were used, the primary probes for each mRNA needed to contain 4 different secondary read sequences in this example. However, in order to reduce the overall length of the primary probes, the pool of targeting sequences for each mRNA target was split at random into two pools. Two secondary read sequences are attached each probe in one of the two pools and the other two secondary read sequences are attached the probes in the other pool. The design criteria for each component are described below.

Primer Design. Specific index primers were generated by a collection of 240,000 published sequences of orthogonal 25-bp long sequences. These sequences were trimmed to 20 bp, selected for a narrow 70 to 80 °C melting temperature, the absence of consecutive repeats of 3 or more base, and the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human genome using BLAST+ (Camacho et al 2009), and primers with 14 or more contiguous bases of homology were eliminated. In a subsequent screening via BLAST+, primers that shared 11 or more contiguous bases or more than 5 bases at the 3' end of any other primer or the T7 promoter were also removed.

Secondary Probes Design. 30-bp long secondary probe sequences were created by concatenating fragments of the orthogonal primer set described above. These secondaries were then screened for orthogonality with other secondaries (no more than 11 basepairs of homology) and for potential off-target binding sites in the human genome (no more than 14 basepairs of homology). Secondary sequences used in this example are provided in Table 1.

Table 1

Bit	Secondary sequences	Sequence number
B1	CGCAACGCTTGGGACGGTTCCAATCGGATC	SEQ ID NO: 2
B2	CGAATGCTCTGGCCTCGAACGAACGATAGC	SEQ ID NO: 3
B3	ACAAATCCGACCAGATCGGACGATCATGGG	SEQ ID NO: 4
B4	CAAGTATGCAGCGCGATTGACCGTCTCGTT	SEQ ID NO: 5
B5	TGCGTCGTCTGGCTAGCACGGCACGCAAAT	SEQ ID NO: 6
B6	AAGTCGTACGCCGATGCGCAGCAATTCACT	SEQ ID NO: 7
B7	CGAAACATCGGCCACGGTCCCGTTGAACTT	SEQ ID NO: 8
B8	ACGAATCCACCGTCCAGCGCGTCAAACAGA	SEQ ID NO: 9
B9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC	SEQ ID NO: 10

B10	GCATGAGTTGCCTGGCGTTGCGACGACTAA	SEQ ID NO: 11
B11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC	SEQ ID NO: 12
B12	GGCCAATGGCCCAGGTCCGTCACGCAATTT	SEQ ID NO: 13
B13	TTGATCGAATCGGAGCGTAGCGGAATCTGC	SEQ ID NO: 14
B14	CGCGCGGATCCGCTTGTCGGGAACGGATAC	SEQ ID NO: 15
B15	GCCTCGATTACGACGGATGTAATTCGGCCG	SEQ ID NO: 16
B16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT	SEQ ID NO: 17

mRNA Targeting Sequence Design. To determine the relative abundance of all the isoforms of all genes expressed in these cell lines, transcriptome profiling data from the ENCODE project for total RNA from A549 and IMR90 cells was processed using the publicly available software *cufflinks*, along with human genome annotations from gencode v18. Gene models corresponding to the most highly expressed isoform were used to build a sequence library in FASTA format recording the dominant isoform of every gene. Genes of interest were selected from this library. These genes were partitioned into 1kb segments, then the software *OligoArray2.1* was used to generate primary probe sequences for the human transcriptome with the following constraints: 30-bp or 40-bp length, depending on the experiment; probe-target melting temperatures greater than 70 °C (variable parameter); no cross hybridization targets with melting temperatures greater than 72 °C (variable parameter); no predicted internal secondary structures with melting temperatures greater than 76 °C (variable parameter); and no single-nucleotide contiguous repeats of 6 or more bases. After OligoArray probe selection, all potential probes that mapped to a different gene were rejected while all potential probes with multiple alignments to the same gene were retained. A BLAST database was assembled from the FASTA library of all expressed genes to screen for probes' uniqueness. For each gene, 14 to 28 targeting sequences produced during the OligoArray processing were selected.

Probe Synthesis – index PCR. The template for specific probe sets were selected from the complex oligopool via limited-cycle PCR. Briefly, 0.5 to1 ng of the complex oligopool was combined with 0.5 micromolar of each primer. The forward primer matched the priming

sequence for the desired subset while the reverse primer was a 5' concatenation of this sequence with a T7 promoter. To avoid the generation of G-quadruplets, which can be difficult to synthesize, the terminal Gs required in the T7 promoter were generated from Gs located at the 5' of the priming region where appropriate. All primers were synthesized by IDT. A 50 microliter reaction volume was amplified either using the KAPA real-time library amplification kit (KAPA Biosystems; KK2701) or via a homemade qPCR mix which included 0.8X EvaGreen (Biotum; 31000-T) and the hot-start Phusion polymerase (New England Biolabs; M0535S). Amplification was followed in real time using Agilent's MX300P or Biorad's CFX Connect. Individual samples were removed immediately before the plateau in amplification to minimize distortion of template abundance due to over-amplification. Individual templates were purified with columns according to the manufacturer's instructions (Zymo DNA Clean and Concentrator; D4003) and eluted in RNase-free deionized water.

Amplification via *in-vitro* transcription. The template was then amplified via *in vitro* transcription. Briefly, 0.5 to 1 microgram of template DNA was amplified into 100-200 microgram of RNA in a single 20-30 microliter reaction with a high yield RNA polymerase (New England Biolabs; E2040S). Reactions were supplemented with 1X RNase inhibitor (Promega RNasin; N2611). Amplification was typically run for 4 to 16 hours at 37 °C to maximize the yield. The RNA was not purified after the reaction and was either stored at -80 °C or immediately converted into DNA as described below.

Reverse Transcription. 1 to 2 nmol of fluorescently-labeled ssDNA probe was created from the above *in vitro* transcription reactions using the reverse transcriptase Maxima H- (Thermo Scientific; EP0751). This enzyme was used because of its higher processivity and temperature resistance, which allowed the conversion of large quantities of RNA into DNA within small volumes at temperatures that disfavor secondary structure formation. The unpurified RNA created above was supplemented with 1.6 mM of each dNTP, 1-2 nmol of fluorescently labeled forward primer, 300 units of Maxima H-, 60 units of RNasin, and a final 1X concentration of the Maxima RT buffer. The final 75 microliter volume was incubated at 50 °C for 60 minutes.

Strand Selection and Purification. The template RNA in the reaction above was then removed from the DNA via alkaline hydrolysis. 75 microliters of 0.25 M EDTA and 0.5 N NaOH were added to each reverse transcription reaction, and the sample was incubated at 95 °C

for 10 minutes. The reaction was immediately neutralized by purifying the ssDNA probe with a modified version of the Zymo Oligo Clean and Concentrator protocol. Specifically, the 5-microgram capacity column was replaced with a 25-microgram or 100-microgram capacity DNA column as appropriate. The remainder of the protocol was run according to the manufacturer's instructions. Probe was eluted in 100 microliter RNase-free deionized water and evaporated in a vacuum concentrator. The final pellet was resuspended in 10 microliters of RNase-free water and stored at -20 °C. Denaturing poly-acrylimide gel electrophoresis and absorption spectroscopy revealed that this protocol typically produced 90-100% incorporation of the fluorescent primer into full length probe and 75-90% recovery of the total fluorescent probe. Thus, without exceeding a 150-microliter reaction volume, this protocol can be used to create ~2 nmol of fluorescent probe.

Cell culture and fixation. A549 and IMR90 cells (American Type Culture Collection) were cultured with Dulbecco's Modified Eagle Medium and Eagle's Minimum Essential Medium respectively. Cells were incubated at 37 °C with 5% CO_2 for 36-48 hours. Cells were fixed in 3% paraformaldehyde (Electron Microscopy Sciences) in PBS for 15 minutes, washed with PBS, and permeabilized in 70% ethanol overnight at 4 °C.

Fluorescence In Situ Hybridization (FISH) – primary (encoding) probes. Cells were hydrated in wash buffers (2xSSC, 50% formamide) for 10 minutes, labelled with primary oligos (0.5 nM per sequence) in hybridization buffers (2xSSC, 50% formamide, 1 mg/mL yeast tRNA, and 10% dextran sulfate) overnight at 37 °C, washed with wash buffers at 47 °C for 10 minutes twice, and washed with 2xSSC twice. Fluorescent fiducial beads (Molecular Probes, F-8809) were added at a 1:10,000 dilution in 2xSSC before imaging.

Secondary probes. Secondary (readout) probes (10 nM) were hybridized in secondary hybridization buffers (2xSSC, 20% formamide, and 10% dextran sulfate) to their primary targets for 30 minutes at 37 °C. Cells remained on the microscope stage during the hybridization. An objective heater was used to maintain the temperature at 37 °C. Cells were washed with secondary wash buffers (2xSSC, 20% formamide).

Fluidics and STORM Imaging. Multiple rounds of sequential labeling, washing, imaging, and bleaching were performed on an automated platform consisted of a fluidics setup and a STORM (stochastic optical reconstruction microscopy) microscope. The fluidics setup included a flow chamber (Bioptech FCS2), a peristaltic pump (Rainin Dynamax RP-1), and three

computer-controlled 8-way valves (Hamilton MVP and Hamilton HVXM 8-5). This system allowed the automated integration of STORM movie collection and secondary hybridization.

The imaging buffer included, 50 mM Tris (pH 8) 10% (w/v) glucose, 1% βME (2-mercaptoethanol) or 25 mM MEA, with or without 2 mM 1,5-cyclooctadiene, and an oxygen scavenging system (0.5 mg/ml glucose oxidase (Sigma-Aldrich) and 40 microgram/ml catalase (Sigma-Aldrich)). A layer of mineral oil was used to seal the imaging buffer, preventing its acidification over the course of multiple hybridizations.

The STORM setup included an Olympus IX-71 inverted microscope configured for oblique incidence excitation. The samples were continuously illuminated with a 642-nm diode-pumped solid-state laser (VFL-P500-642; MPB communications). A 405-nm solid-state laser (Cube 405-100C; Coherent) was used for activation of dyes. Fluorescence was collected using an Olympus (UPlanSApo 100x, 1.4 NA) objective lens and passed through a custom dichroic, as well as a quad-view beam splitter. All movies were recorded using an EMCCD camera (Andor iAxon 897), imaging at 60Hz. The 512x256 field of view of the camera was split into separate 256x256 pixel movies prior to saving. The left half of this field of view contained the STORM data and the right half contained images of the fluorescent feducial beads. These latter movies were downsampled to 1 Hz prior to saving. During data acquisition, a home-built focus lock was used to maintain a constant focal plane. STORM movies included 20,000 to 30,000 frames in STORM buffer while the bleach movies included 10,000 frames in wash buffer.

Image Analysis – analysis of single-molecule localizations. Movies of single molecule localizations and fluorescent feducial beads were processed separately using a previously published single-emitter localization software.

Image Registration. The starting position of the beads from each round of hybridization were used to align movies from each round. The 2D autocorrelation between bead images of consecutive hybridizations followed by nearest-neighbor matching was used to match beads between images. The pair of beads with the most similar displacement vector were used to compute a rigid translation-rotation warp to align the beads. This alignment method is robust to samples in which multiple feducials are displaced or come detached and reattach during imaging.

Drift Correction. Drift during image acquisition was corrected using the trajectory of the feducial beads (recorded at 1 Hz). Bead positions were linked in each frame. The trajectory of the two beads that moved in the most correlated fashion was taken as the drift trajectory.

mRNA Cluster Calling. Localizations were first screened to be above a threshold number of photons (generally 2000) and required to be within 32 nm of 5 other localizations (parameters may be tuned). The remaining molecule localizations were binned in a 2D histogram of 10x10 nm bins (bin size is a variable parameter). All connected bins were taken to be part of a cluster (diagonal contacts are classified as connected). Clusters were required to have more than 80 total localizations across all hybridizations (variable parameter) to be called an mRNA cluster. The weighted centroids of these clusters from the 2D histogram were recorded as the mRNA positions.

A given cluster is recorded to be represented in an individual hybridization round if more than 9 localizations (variable parameter) are found within a 48 nm radius (variable parameter) of the centroid for that mRNA in each hybridization round.

Cluster Decoding. For each mRNA cluster, a codeword is readout, including "0"s for all the hybridization rounds in which less than the threshold number of localizations are found near the centroid and "1" s for the rounds where above threshold number of localizations are counted. The SECDED codebook decoded these as either perfect matches to target mRNA codewords, correctable errors which can be unambiguously mapped back to target mRNA, or uncorrectable errors, which differed by two or more letters from the words in the codebook.

Fig. 2A shows a STORM image of a cell. Fig. 2B shows a zoom in of the boxed region in Fig. 2A. Each dot indicates a localization. Localizations from different rounds of imaging are shown differently. Fig. 2C shows a representative cluster of localizations from the boxed region in Fig. 2B. The cluster shows localization signals from 4 different hybridizations. This cluster is a putative mRNA encoded with codeword [0 1 0 1 1 1 0 0]. Fig. 2D shows a reconstructed cell image of 14 genes after decoding and error correction. Each gene is shown differently. Fig. 2E shows measured gene expression for the 14 genes from the cell. Fig. 2F shows a comparison of transcript count with ensemble RNA-sequencing data. Fig. 2G shows correlation of transcript expression level between two cells detected using the described approach.

EXAMPLE 4

The following examples are generally directed to multiplexed single-molecule imaging with error-robust encoding allowing for simultaneous measurements of thousands of RNA species in single cells. In general, knowledge of the expression profile and spatial landscape of RNAs in individual cells is essential for understanding the rich repertoire of cellular behaviors.

The following examples reports various techniques directed to single-molecule imaging approaches that allow the copy numbers and spatial localizations of thousands of RNA species to be determined in single cells. Some of these techniques are called Multiplexed Error-Robust Fluorescence in Situ Hybridization or "MERFISH."

Using error-robust encoding schemes to combat single-molecule labeling and detection errors, these examples demonstrated the imaging of hundreds to thousands of unique RNA species in hundreds of individual cells. Correlation analysis of the ~10^4 to ~10^6 pairs of genes allowed constraints on gene regulatory networks, prediction of novel functions for many unannotated genes, and identification of distinct spatial distribution patterns of RNAs that correlate with properties of the encoded proteins.

System-wide analyses of the abundance and spatial organization of RNAs in single cells promise to transform understanding in many areas of cell and developmental biology, such as the mechanisms of gene regulation, the heterogeneous behavior of cells, and the development and maintenance of cell fate. Single-molecule fluorescence *in situ* hybridization (smFISH) has emerged as a powerful tool for studying the copy number and spatial organization of RNAs in single cells either in isolation or in their native tissue context. Taking advantage of its ability to map the spatial distributions of specific RNAs with high resolution, smFISH has revealed the importance of subcellular RNA localization in diverse processes such as cell migration, development, and polarization. In parallel, the ability of smFISH to precisely measure the copy numbers of specific RNAs without amplification bias has allowed quantitative measurement of the natural fluctuations in gene expression, which has in turn elucidated the regulatory mechanisms that shape such fluctuations and their role in a variety of biological processes.

However, application of the smFISH approach to many systems-level questions remains limited by the number of RNA species that can be simultaneously measured in single cells. State-of-the-art efforts using combinatorial labeling by either color-based barcodes or sequential hybridization have enabled simultaneous measurements of 10-30 different RNA species in individual cells, yet many interesting biological questions would benefit from the measurement of hundreds to thousands of RNAs within a single cell, which are not achievable using such techniques. For example, analysis of how the expression profile of such a large number of RNAs vary from cell to cell and how these variations correlate among different genes could be used to systematically identify co-regulated genes and map regulatory networks; knowledge of the

subcellular organizations of numerous RNAs and their correlations could help elucidate molecular mechanisms underlying the establishment and maintenance of many local cellular structures; and RNA profiling of individual cells in native tissues could allow *in situ* identification of cell type.

The following examples generally discuss certain techniques called MERFISH, which are highly multiplexed smFISH imaging methods that substantially increase the number of RNA species that can be simultaneously imaged in single cells by using combinatorial labeling and sequential imaging with error-robust encoding schemes. These examples demonstrate this multiplexed imaging approach by simultaneously measuring 140 RNA species using an encoding scheme that can both detect and correct errors and 1001 RNA species using an encoding scheme that can detect but not correct errors. It should be understood that these numbers are by way of exemplification only, not limitation. Correlation analyses of the copy number variations and spatial distributions of these genes allowed us to identify groups of genes that are co-regulated and groups of genes that share similar spatial distribution patterns inside the cell.

Combinatorial labeling with error-robust encoding schemes. Combinatorial labeling that identifies each RNA species by multiple (N) distinct signals offers a route to rapidly increase the number of RNA species that can be probed simultaneously in individual cells (Fig. 5A). However, this approach to scaling up the throughput of smFISH to the systems scale faces a significant challenge because not only does the number of addressable RNA species increases exponentially with N, but the detection error rates also grow exponentially with N (Figs. 5B-5D). Imagine a conceptually simple scheme to implement combinatorial labeling, where each RNA species is encoded with a N-bit binary word and the sample is probed with N corresponding rounds of hybridization, each round targeting only the subset of RNAs that should read '1' in the corresponding bit (Fig. 11). N rounds of hybridization would allow 2^N-1 RNA species to be probed. With just 16 hybridizations, over 64,000 RNA species, which should cover the entire human transcriptome including both messenger RNAs (mRNAs) and non-coding RNAs, could be identified (Fig. 5B; upper symbols). However, as N increases, the fraction of RNAs properly detected (the calling rate) would rapidly decrease and, more troublingly, the fraction of RNAs that are identified as incorrect species (the misidentification rate) would rapidly increase (Fig. 5C, lower symbols; Fig. 5D, upper symbols). With realistic error rates per hybridization

(measured below), the majority of RNA molecules would be misidentified after 16 rounds of hybridizations!

To address this challenge, error-robust encoding schemes were designed, in which only a subset of the 2^N-1 words separated by a certain Hamming distance were used to encode RNAs. In a codebook where the minimum Hamming distance is 4 (HD4 code), at least four bits must be read incorrectly to change one code word into another (Fig. 12A). As a result, every single-bit error produces a word that is uniquely close to a single code word, allowing such errors to be detected and corrected (Fig. 12B). Double-bit errors produce words with an equal Hamming distance of 2 from multiple code words and, thus, can be detected but not corrected (Fig. 12C). Such a code should substantially increase the calling rate and reduce the misidentification rate (Fig. 5C and 5D, middle symbols). To further account for the fact that it is more likely to miss a hybridization event (an 1-->0 error) than to misidentify a background spot as an RNA (an 0-->1 error) in smFISH measurements, a modified HD4 (MHD4) code was designed, in which the number of '1' bits were kept both constant and relatively low, only four per word, to reduce error and avoid biased detection. This MHD4 code should further increase the calling rate and reduce the misidentification rate (Fig. 5C, upper symbols; Fig. 5D, lower symbols).

In addition to the error considerations, several practical challenges have also made it difficult to probe a large number of RNA species, such as the high cost of the massive number of fluorescently labeled FISH probes needed and the long time required to complete many rounds of hybridization. To overcome these challenges, in this example, a two-step labeling scheme was designed to encode and readout cellular RNAs (Fig. 5E). First, cellular RNAs were labeled with a set of encoding probes (also called primary probes), each probe comprising a RNA targeting sequence and two flanking readout sequences. Four of the N unique readout sequences were assigned to each RNA species based on the MHD4 code word of the RNA. Second, these N readout sequences were identified with complementary FISH probes, the readout probes (also called secondary probes) via N rounds of hybridization and imaging, each round using a unique readout probe. To increase the signal to background ratio, every cellular RNA was labeled with ~192 encoding probes. Because each encoding probe contained two of the four readout sequences associated with that RNA (Fig. 5E), a maximum of ~96 readout probes can bind to each cellular RNA per hybridization round. To generate the massive number of encoding probes required, they were amplified from array-derived oligonucleotide pools containing tens of

thousands of custom sequences using an enzymatic amplification process comprising *in vitro* transcription followed by reverse transcription (Fig. 13, see below regarding probe synthesis). This two-step labeling approach significantly diminished the total hybridization time for an experiment: it was found that efficient hybridization to the readout sequences took only 15 minutes whereas efficient direct hybridization to cellular RNA required more than 10 hours.

Fig. 5 describes MERFISH, a highly multiplexed smFISH approach using combinatorial labeling and error-robust encoding. Fig. 5A shows a schematic depiction of the identification of multiple RNA species in N rounds of imaging. Each RNA species is encoded with a N-bit binary word and during each round of imaging, only the subset of RNAs that should read '1' in the corresponding bit emit signal. Figs. 5B-5D show the number of addressable RNA species (Fig. 5B), the rate at which these RNAs are properly identified (calling rate) (Fig. 5C), and the rate at which RNAs are incorrectly identified as a different RNA species (misidentification rate) (Fig. 5D) plotted as a function of the number of bits (N) in the binary words encoding RNA. In Figs. 5B and 5D, the upper dots are a simple binary code that includes all 2^N-1 possible binary words; the middle dots are the HD4 code where the Hamming distance separating words is 4; and the lower dots are the modified HD4 (MHD4) code where the number of '1' bits are kept at four. These are reversed in Fig. 5C.

The calling and misidentification rates are calculated with per bit error rates of 10% for the 1-->0 error and 4% for the 0-->1 error. Fig. 5E is as schematic diagram of the implementation of a MHD4 code for RNA identification. Each RNA species is first labeled with ~192 encoding probes that convert the RNA into a unique combination of readout sequences (Encoding hyb). These encoding probes each contain a central RNA targeting region flanked by two readout sequences, drawn from a pool of N different sequences, each associated with a specific hybridization round. Encoding probes for a specific RNA species contain a unique combination of four of the N readout sequences, which correspond to the four hybridization rounds where this RNA should read '1'. N subsequent rounds of hybridization with the fluorescent readout probes were used to probe the readout sequences (hyb 1, hyb 2, ..., hyb N). The bound probes were inactivated by photobleaching between successive rounds of hybridization. For clarity, only one possible pairing of the readout sequences is depicted here for the encoding probes; however, all possible pairs of the four readout sequences are used at the same frequency and distributed randomly along each cellular RNA in the actual experiments.

Fig. 11 shows a schematic description of a combinatorial labeling approach based on a simple binary code. In a conceptually simple labeling approach, 2^N-*1* different RNA species can be uniquely encoded with all *N*-bit binary words (excluding the word with all '0's). In each hybridization round, FISH probes that are targeted to all RNA species that have a '1' in the corresponding bit are included. To increase the ability to discriminate RNA spots from background, each RNA is addressed with multiple FISH probes per hybridization round. Signal from the bound probes is extinguished before the next round of hybridization. This process continues for all *N* hybridization rounds (hyb 1, hyb 2, ...), and all 2^N-*1* RNA species can be identified by the unique on-off pattern of fluorescence signals in each hybridization round.

Fig. 12 shows schematic descriptions of Hamming distance and its use in the identification and correction of errors. Fig. 12A is a schematic representation of a Hamming distance of 4. Figs. 12B and 12C are schematics showing the ability of an encoding scheme with Hamming distance 4 to correct single-bit errors (Fig. 12B) or detect but not correct double-bit errors (Fig. 12C). Arrows highlight bits at which the indicated words differ. Two code words are separated by a Hamming distance of 4 if one of the words has to flip four bits from '1' to '0' or '0' to '1' to convert into the other word. Single-bit error correction is possible because if a measured word differs from a legitimate code word by only one bit, it is most likely an error that arises from misreading this code word, since the code words of all the other RNA species will differ from the measured word by at least three bits. In this case, the measured word can be corrected to a code word that differs by only one bit. If a measured word differs from a legitimate code word by two bits, this measured word can still be identified as an error, but correction is no longer possible since more than one legitimate code word differs from this measured word by two bits.

Fig. 13 shows production of the library of encoding probes. An array-synthesized complex oligopool, containing ~100k sequences, is used as a template for the enzymatic amplification of the encoding probes for different experiments. Each template sequence in the oligopool contains a central target region that can bind to a cellular RNA, two flanking readout sequences, and two flanking index primers. In the first step, the required template molecules for a specific experiment are selected and amplified with an indexed PCR reaction. To allow amplification via *in vitro* transcription, a T7 promotor is added to the PCR products during this step. In the second step, RNA is amplified from these template molecules via *in vitro*

transcription. In the third step, this RNA is reverse transcribed back into DNA. In the final step, the template RNA is removed via alkaline hydrolysis, leaving only the desired ssDNA probes. This protocol produces ~2 nmol of complex pools of encoding probes containing ~20,000 different sequences for the 140-gene experiments or ~100,000 different sequences for the 1001-gene experiments.

EXAMPLE 5

This example illustrates the measurement of 140 genes with MERFISH using a 16-bit MHD4 Code. To test the feasibility of this error-robust, multiplexed imaging approach, this example uses a 140-gene measurement on human fibroblast cells (IMR90) using a 16-bit MHD4 code to encode 130 RNA species while leaving 10 code words as misidentification controls (Fig. 20). After each round of hybridization with the fluorescent readout probes, cells were imaged by conventional wide-field imaging with an oblique-incidence illumination geometry. Fluorescent spots corresponding to individual RNAs were clearly detected and were then efficiently extinguished via a brief photobleaching step (Fig. 6A). The sample was stable throughout the 16 rounds of iterative labeling and imaging. The change in the number of fluorescent spots from round to round matched the predicted change based on the relative abundances of RNA species targeted in each round derived from bulk sequencing, and a systematic decreasing trend with increasing number of hybridization rounds was not observed (Fig. 14A). The average brightness of the spots varied from round to round with a standard deviation of 40%, likely due to different binding efficiencies of the readout probes to the different readout sequences on the encoding probes (Fig. 14B). Only a small, systematic decreasing trend in the spot brightness with increasing hybridization rounds was observed, which was on average 4% per round (Fig. 14B).

Next, binary words were constructed from the observed fluorescent spots based on their on-off patterns across the 16 hybridization rounds (Fig. 6B-6D). If the word exactly matched one of the 140 MHD4 code words (exact matches) or differed by only one bit (error-correctable matches), it was assigned to the corresponding RNA species (Fig. 6D). Within the single cell depicted in Figs. 6A and 6B, more than 1500 RNA molecules corresponding to 87% of the 130 encoded RNA species were detected after error correction (Fig. 6E). Similar observations were made in ~400 cells from 7 independent experiments. On average, ~4 times as many RNA molecules and ~2 times as many RNA species were detected per cell after error correction as compared with the values obtained before error correction (Fig. 15).

Two types of errors can occur in the copy number measurement of each RNA species: 1) Some molecules of this RNA species are not detected, leading to a drop in calling rate, and 2) some molecules from other RNA species are misidentified as this RNA species. To assess the extent of misidentification, the 10 misidentification control words were utilized, i.e., code words that were not associated with any cellular RNA. Although matches to these control words were observed, they occurred far less frequently than the real RNA-encoding words: 95% of the 130 RNA-encoding words were counted more frequently than the median count for these control words. Moreover, it was typically found that the ratio of the number of exact matches to the number of matches with one-bit errors for a real RNA-encoding word was substantially higher than the same ratios observed for the misidentification controls, as expected (Figs. 16A and 16B). Using this ratio as a measure of the confidence in RNA identification, it was found that 91% of the 130 RNA species had a confidence ratio greater than the maximum confidence ratio observed for the misidentification controls (Fig. 6F), demonstrating a high accuracy of RNA identification. Subsequent analyses were conducted only on these 91% of genes.

To estimate the calling rate, the error-correction ability of the MHD4 code was utilized to determine the 1-->0 error rates (10% on average) and 0-->1 error rates (4% on average) for each hybridization round (Figs. 16C and 16D). Using these error rates, an ~80% calling rate for individual RNA species after error correction was estimated, i.e. ~80% of the fluorescent spots corresponding to a RNA species were decoded correctly (Fig. 16E). It is noted that although the remaining 20% of spots contributed to a loss in detection efficiency, most of them did not cause species misidentification because they were decoded as double-bit error words and discarded.

To test for potential technical bias in these measurements, the same 130 RNAs species were probed with a different MHD4 codebook by shuffling the code words among different RNA species (Fig. 20) and changing the encoding probe sequences. Measurements with this alternative code gave similar misidentification and calling rates (Fig. 17). The copy numbers of individual RNA species per cell measured with these two codebooks showed excellent agreement with a Pearson correlation coefficient of 0.94 (Fig. 6G), indicating that the choice of encoding scheme did not bias the measured counts.

In order to validate the copy numbers derived from the MERFISH experiments, conventional smFISH measurements were performed on 15 of the 130 genes, selected from the full measured abundance range of three orders of magnitude. For each of these genes, both the

average copy number and the copy number distribution across many cells agreed quantitatively between the MERFISH and conventional smFISH measurements (Fig. 18A and 18B). The ratio of the copy numbers determined by these two approaches was 0.82 +/- 0.06 (mean +/- SEM across the 15 measured RNA species, Fig. 18B), which agreed with the estimated 80% calling rate for the multiplexed imaging approach. The quantitative match between this ratio and the estimated calling rate over the full measured abundance range additionally supports the assessment that the misidentification error was low. Given that the agreement between the MERFISH and conventional smFISH results extended to the genes at the lowest measured abundance (<1 copy per cell, Fig. 18B), it was estimated that the measurement sensitivity was at least 1 copy per cell.

As a final validation, the abundance of each RNA species averaged over hundreds of cells was compared to those obtained from a bulk RNA sequencing measurement that were performed on the same cell line. The imaging results correlated remarkably well with bulk sequencing results with a Pearson correlation coefficient of 0.89 (Fig. 6H).

Fig. 6 shows simultaneous measurement of 140 RNA species in single cells using MERFISH with a 16-bit MHD4 code. Fig. 6A shows images of RNA molecules in an IMR90 cell after each hybridization round (hyb 1 – hyb 16). The image after photobleaching (bleach 1) demonstrated efficient removal of fluorescent signals between hybridizations. Fig. 6B shows the localizations of all detected single molecules in this cell colored based on their measured binary words. Inset: the composite fluorescent image of the 16 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. A circle indicates an unidentifiable molecule, the binary word of which does not match any of the 16-bit MHD4 code words even after error correction. Fig. 6C shows fluorescent images from each round of hybridization for the boxed sub-region in Fig. 6B with circles indicating potential RNA molecules. Fig. 6D shows corresponding words for the spots identified in Fig. 6C. Crosses represent the corrected bits. Fig. 6E shows the RNA copy number for each gene observed without (lower) or with (higher) error correction in this cell. Fig. 6F shows the confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words (right) normalized to the maximum value observed from the misidentification controls (dashed line). Fig. 6G is a scatter plot of the average copy number of each RNA species per cell measured with two shuffled codebooks of the MHD4 code. The Pearson correlation coefficient is 0.94 with a p-

value of $1x10^{-53}$. The dashed line corresponds to the $y = x$ line. Fig. 6H is a scatter plot of the average copy number of each RNA species per cell versus the abundance determined by bulk sequencing in fragments per kilobase per million reads (FPKM). The Pearson correlation coefficient between the logarithmic abundances of the two measurements was 0.89 with a p-value of $3x10^{-39}$.

Fig. 14 shows the number and average brightness of the fluorescent spots detected in the 16 rounds of hybridization before and after photobleaching. Fig. 14A shows the number of fluorescent spots observed per cell before (higher) and after (lower) photobleaching as a function of hybridization round averaged across all measurements with the first 16-bit MHD4 code. Photobleaching reduces the number of fluorescent spots by two or more orders of magnitude. Hybridization rounds without lower bars represent rounds in which no molecules were observed after bleaching. Also depicted is the expected change in the number of fluorescent spots from round to round (circles) predicted based on the relative abundances of the RNA species targeted in each hybridization round derived from bulk RNA sequencing. The average discrepancy between the observed and predicted number of spots for each hybridization is only 15% of the mean number of spots. This discrepancy does not systematically increase with the number of hybridization rounds. Fig. 14B shows the average brightness of the identified fluorescent spots in each hybridization round averaged across all measurements with the first 16-bit MHD4 code both before (upper) and after (lower) photobleaching. Brightness varies by 40% (standard deviation) across different hybridization rounds. The variation pattern is reproducible between experiments with the same code, likely due to differences in the binding efficiency of the readout probes to the different readout sequences. There is a small systematic trend of decrease in the brightness with increasing hybridization rounds, which is on average 4% per round. Photobleaching extinguishes fluorescence to a level similar to that of the autofluorescence of the cell.

Fig. 15 shows error correction substantially increases the numbers of RNA molecules and RNA species detected in individual cells. Fig. 15A shows a histogram of the ratio of the total number of molecules detected per cell with error correction to the number measured without error correction. Fig. 15B is a histogram of the total number of RNA species detected in each cell with error correction to that without error correction. Both ratios are determined for ~200 cells and the histograms are constructed from these ratios.

Fig. 16 shows characterization of the misidentification and calling rates of RNA species for the 140-gene experiments using a specific 16-bit MHD4 code. Fig. 16A shows the number of measured words exactly matching the code word corresponding to FLNA, represented by the bar in the center of the circle, and the number of measured words with one-bit error compared to the code word of FLNA, represented by the 16 bars on the circle. Fig. 16B is the same as Fig. 16A, but for a code word that was not assigned to any RNA, i.e., a misidentification control word. The solid lines connect the exact match to one-bit error words that are generated by 1-->0 errors. Based on the observation that the ratio of the number of exact matches to the number of error-correctable matches for a real RNA-encoding word was typically substantially higher than the same ratios observed for the misidentification controls, this ratio was defined as a confidence ratio for RNA identification. The confidence ratio measured for all 130 RNA species (center bar) and 10 misidentification control words not assigned to any RNA (outer bars) using this 16-bit MHD4 code is show in Fig. 6F. Fig. 16C and 16D show the average error rates for the 1-->0 error (Fig. 16C) and 0-->1 error (Fig. 16D) for each hybridization round. Fig. 16E shows the calling rate for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance, which spans three orders of magnitude. The calling rates are largely independent of the abundance of the gene.

Fig. 17 shows characterization of the misidentification and calling rates for a second 16-bit MHD4 code. In this second encoding scheme, the 140 code words were shuffled among different RNA species and changed the encoding probe sequences. Fig. 17A shows the normalized confidence ratio measured for the 130 RNA species (left) and the 10 misidentification control words not assigned to any RNA (right). The normalized confidence ratio is determined the same way as in Fig. 6F. Figs. 17B and 17C show the average error rates determined for the 1-->0 error (Fig. 17B) and 0-->1 error (Fig. 17C) for each hybridization round. Fig. 17D shows the calling rate determined for each RNA species estimated from the 1-->0 and 0-->1 error rates. Genes are sorted from left to right based on the measured abundance.

Fig. 18 shows a comparison of the MERFISH measurements with conventional smFISH results for a subset of genes. Fig. 18A shows the distributions of RNA copy numbers in single cells for three example genes KIAA1199, DYNC1H1, and LMTK2 in the high, medium, and low abundance ranges, respectively. Lighter bars: distributions constructed from ~400 cells in the 140-gene measurements using the MHD4 codes. Darker bars: distributions constructed from

~100 cells in the conventional smFISH measurements. Fig. 18B shows a comparison of the average RNA copy numbers per cell measured in the 140-gene experiments using the MHD4 codes to those determined by conventional smFISH for 15 genes. The average ratio of the copy number measured using the MHD4 measurements to that measured using conventional smFISH was 0.82 +/- 0.06 (mean +/- SEM across 15 genes). The dashed line corresponds to the $y = x$ line.

Fig. 20 shows two different codebooks for the 140-gene experiments. The specific code words of the 16-bit MHD4 code assigned to each RNA species studied in the two shuffles of the 140-gene experiment. The "Genes" columns contain the name of the gene. The "Codewords" columns contain the specific binary word assigned to each gene.

EXAMPLE 6

This example is generally directed to high-throughput analysis of cell-to-cell variation in gene expression. The MERFISH approach allows parallelization of measurements of many individual RNA species and co-variation analysis between different RNA species. In this example, the parallelization aspect was first illustrated by examining the cell-to-cell variation in the expression level of each of the measured genes (Fig. 7A). To quantify the measured variation, the Fano factors, defined as the ratio of the variance to the mean RNA copy number, were computed for all measured RNA species. The Fano factors substantially deviated from 1, the value expected for a simple Poisson process, for many genes and exhibited an increasing trend with the mean RNA abundance (Fig. 7B). This trend of increasing Fano factors with mean RNA abundance can be explained by changes in the transcription rate and/or promoter off-switching rates but not by changes in the promoter on-switching rate.

Moreover, several RNA species were identified with substantially larger Fano factors than this average trend. For example, it was found that SLC5A3, CENPF, MKI67, TNC and KIAA1199 displayed Fano factor values substantially higher than those of the other genes expressed at similar abundance levels. The high variability of some of these genes can be explained by their association with the cell cycle. For example, two of these particularly 'noisy' genes MKI67 and CENPF were both annotated as cell-cycle related genes, and based on their bimodal expression (Fig. 7C), it is proposed that their transcription is strongly regulated by the

cell cycle. Other high-variability genes did not show the same bimodal expression patterns and are not known to be associated with the cell cycle.

Analysis of co-variations in the expression levels of different genes can reveal which genes are co-regulated and elucidate gene regulatory pathways. At the population level, such analysis often requires the application of external stimuli to drive gene expression variation; hence, correlated expression changes can be observed among genes that share common regulatory elements influenced by the stimuli. At the single-cell level, one can take advantage of the natural stochastic fluctuations in gene expression for such analysis and can thus study multiple regulatory networks without having to stimulate each of them individually. Such co-variation analysis can constrain regulatory networks, suggest new regulatory pathways, and predict function for unannotated genes based on associations with co-varying genes.

This approach was applied to the 140-gene measurements and the ~10,000 pairwise correlation coefficients describing how the expression levels of each pair of genes co-varied from cell to cell were examined. Many of the highly variable genes showed tightly correlated or anti-correlated variations (Fig. 7C). To better understand the correlations for all gene pairs, a hierarchical clustering approach was adopted to organize these genes based on their correlation coefficients (Fig. 7D). From the cluster tree structure, seven groups of genes with substantially correlated expression patterns were identified (Fig. 7D). Within each of the seven groups, every gene showed significantly stronger average correlation with other members of the group than with genes outside the group. To further validate and understand these groups, gene ontology (GO) terms enriched in each of these seven groups were identified. Notably, the enriched GO terms within each group shared similar functions and were largely unique to each group (Fig. 7E), validating the notion that the observed co-variation in expression reflects some commonalities in the regulation of these genes.

This example describes two of these groups as illustrative examples. The predominant GO terms associated with Group 1 were terms associated with the extracellular matrix (ECM) (Figs. 7D and 7E). Notable members of this group included ECM components, such as FBN1, FBN2, COL5A, COL7A and TNC, and glycoproteins linking the ECM and cell membranes, such as VCAN and THBS1. The group also included an unannotated gene, KIAA1199, which may predict to play a role in ECM metabolism based on its association with this cluster. Indeed,

this gene has recently been identified as an enzyme involved in the regulation of hyaluronan, a major sugar component of the ECM.

Group 6 contained many genes that encode vesicle transport proteins and proteins associated with cell motility (Figs. 7D and 7E). The vesicle transport genes included microtubule motors and related genes DYNC1H, CKAP1, and factors associated with vesicle formation and trafficking, like DNAJC13 and RAB3B. Again, an unannotated gene, KIAA1462, was found within this cluster. Based on its strong correlation with DYNC1H1 and DNAJC13, it is predicted that this gene may be involved in vesicle transport. The cell motility genes in this group included actin-binding proteins like AFAP1, SPTAN1, SPTBN1, and MYH10, and genes involved in the formation of adhesion complexes, like FLNA and FLNC. Several GTPase-associated factors involved in the regulation of cell motility, attachment and contraction also fell into this group, including DOCK7, ROCK2, IQGAP1, PRKCA, and AMOTL1. The observation that some cell motility genes correlated with vesicle transport genes is consistent with the role of vesicle transport in cell migration. An additional interesting feature of group 6 is that a subset of these genes, in particular those related to cell motility, were anti-correlated with members of the ECM group discussed above (Fig. 7D). This anti-correlation may reflect regulatory interactions that mediate switching of cells between adherent and migratory states.

Fig. 7 shows cell-to-cell variations and pairwise correlations for the RNA species determined from the 140 gene measurements. Fig. 7A shows a comparison of gene expression levels in two individual cells. Fig. 7B shows Fano factors for individual genes. Error bars represent standard error of the mean determined from 7 independent data sets. Fig. 7C shows Z-scores of the expression variations of four example pairs of genes showing correlated (top two) or anti-correlated (bottom two) variation for 100 randomly selected cells. The Z-score is defined as the difference from the mean normalized by the standard deviation. Fig. 7D is a matrix of the pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes, shown together with the hierarchical clustering tree. The seven groups identified by a specific threshold on the cluster tree (dashed line) are indicated by the black boxes in the matrix and lines on the tree, with grey lines on the tree indicating ungrouped genes. Different threshold choices on the cluster tree could be made to select either smaller subgroups with tighter correlations or larger super-groups containing more weakly coupled subgroups. Two of the seven groups are enlarged on the right. Fig. 7E shows enrichment of 30 selected, statistically

significantly enriched GO terms in the seven groups. Enrichment refers to the ratio of the fraction of genes within a group that have the specific GO term to the fraction of all measured genes having that term. Not all of the GO terms presented here are in the top 10 list.

EXAMPLE 7

This example illustrates mapping spatial distributions of RNAs. As an imaging based approach, MERFISH also allowed the investigation of the spatial distributions of many RNA species simultaneously. Several patterns emerged from the visual inspection of individual genes, with some RNA transcripts enriched in the perinuclear region, some enriched in the cell periphery, and some scattered throughout the cell (Fig. 8A). To identify genes with similar spatial distributions, the correlation coefficients for the spatial density profiles of all pairs of RNA species were determined, and these RNAs were organized based on the pairwise correlations again using the hierarchical clustering approach. The correlation coefficient matrix showed groups of genes with correlated spatial organizations, and the two most notable groups with the strongest correlations are indicated in Fig. 8B. Group I RNAs appeared enriched in the perinuclear region whereas group II RNAs appeared enriched near the cell peripheral region (Fig. 8C). Quantitative analysis of the distances between each RNA molecule and the cell nucleus or the cell periphery indeed confirmed this visual impression (Fig. 8D).

Group I contained genes encoding extracellular proteins such as FBN1, FBN2 and THSB1, secreted proteins such as PAPPA, and integral membrane proteins such as LRP1 and GPR107. These proteins have no obvious commonalities in function. Rather a GO analysis showed significant enrichment for location terms, such as extracellular region, basement membrane, or perivitelline space (Fig. 8E). To reach these locations, proteins must pass through the secretion pathway, which often requires translation of mRNA at the endoplasmic reticulum (ER). Thus, it is proposed that the spatial pattern that were observed for these mRNAs reflects their co-translational enrichment at the ER. The enrichment of these mRNAs in the perinuclear region (Figs. 8C and 8D, lighter shading), where the rough ER resides, supports this conclusion.

Group II contained genes encoding the actin-binding proteins, including filamins FLNA and FLNC, talin TLN1, and spectrins SPTAN1 and SPTBN1; the microtubule-binding protein CKAP5; and the motor proteins MYH10 and DYNC1H1. This group was enriched with GO terms such as cortical actin cytoskeleton, actin filament binding, and cell-cell adherens junction (Fig. 8E). Beta-actin mRNA may be enriched near the cell periphery in fibroblasts as are

mRNAs that encode members of the actin-binding Arp2/3 complex. The enrichment of group II mRNAs in the peripheral region of the cells (Figs. 8C and 8D) suggests that the spatial distribution of the Group II genes might be related to the distribution of actin cytoskeleton mRNAs.

Fig. 8 shows distinct spatial distributions of RNAs observed in the 140-gene measurements. Fig. 8A shows examples of the spatial distributions observed for four different RNA species in a cell. Fig. 8B is a matrix of the pairwise correlation coefficients describing the degree with which the spatial distributions of each gene pair is correlated, shown together with the hierarchical clustering tree. Two strongly correlating groups are indicated by the black boxes on the matrix and shading on the tree. Fig. 8C shows the spatial distributions of all RNAs in the two groups in two example cells. Lighter symbols: group I genes; darker symbols: group II genes. Fig. 8D shows average distances for genes in group I and genes in group II to the cell edge or the nucleus normalized to the average distances for all genes. Error bars represent SEM across 7 data sets. Fig. 8E shows enrichment of GO terms in each of the two groups.

EXAMPLE 8

This example illustrates measuring 1001 genes with a 14-bit MHD2 code. This example further increases the throughput of MERFISH measurements by simultaneously imaging ~1000 RNA species. This increase could be achieved with the MHD4 code by increasing the number of bits per code word to 32 while maintaining the number of '1' bits per word at four (Fig. 5B). While the stability of the samples across many hybridization rounds (Fig. 14) suggests that such an extension is potentially feasible, an alternative approach is shown here that did not require an increase in the number of hybridizations by relaxing the error correction requirement but keeping the error detection capability. For example, by reducing the Hamming distance from 4 to 2, all 14-bit words could be used that contain four '1' bits to encode 1001 genes and these RNAs were probed with only 14 rounds of hybridization. However, because a single error can produce a word equally close to two different code words, error correction is no longer possible for this modified Hamming-distance-2 (MHD2) code. Hence, it was expected that the calling rate would be lower and the misidentification rate to be higher with this encoding scheme.

To evaluate the performance of this 14-bit MHD2 code, 16 of the 1001 possible code words were set aside as misidentification controls and used the remaining 985 words to encode cellular RNAs. Among these 985 RNAs included 107 RNA species probed in the 140-gene

experiments as an additional control. The 1001-gene experiments were performed in IMR90 cells using a similar procedure as described above. To allow all encoding probes to be synthesized from a single 100,000-member oligopool, the number of encoding probes per RNA species was reduced to ~94. Fluorescent spots corresponding to individual RNA molecules were again detected in each round of hybridization with the readout probes and, based on their on-off patterns, these spots were decoded into RNA (Figs. 9A, 19A and 19B). 430 RNA species were detected in the cell shown in Fig. 9A, and similar results were obtained in ~200 imaged cells in 3 independent experiments.

As expected, the misidentification rate of this scheme was higher than that of the MHD4 code. 77% of all real RNA words were detected more frequently than the median count for the misidentification controls instead of the 95% value observed in the MHD4 measurements. Using the same confidence ratio analysis as described above, it was found that 73% (instead of 91% for the MHD4 measurements) of the 985 RNA species were measured with a confidence ratio larger than the maximum value observed for the misidentification controls (Fig. 19C). RNA copy numbers measured from these 73% RNA species showed excellent correlation with the bulk RNA sequencing results (Pearson correlation coefficient $r = 0.76$; Fig. 9B, black). It is worth noting that the remaining 27% of the genes still exhibit good, albeit lower, correlation with the bulk RNA sequencing data ($r = 0.65$; Fig. 9B, red), but the conservative measure of excluding them from further analysis was taken.

The lack of an error correction capability also decreased the calling rate of each RNA species: When comparing the 107 RNA species common in both the 1001-gene and 140-gene measurements, it was found that the copy numbers per cell of these RNA species were lower in the 1001-gene measurements (Figs. 9C and 19D). The total count of these RNAs per cell was ~1/3 of that observed in the 140-gene measurements. Thus the lack of error correction in the MHD2 code produced a ~3-fold decrease in the calling rate, which is consistent with the ~4-fold decrease in calling rate observed for the MHD4 code when error correction was not applied. As expected from the quantitative agreement between 140-gene measurements and conventional smFISH results, comparison of the 1001-gene measurements with conventional smFISH results for 10 RNA species also indicated a ~3-fold drop in calling rate (Fig. 18C). Despite the expected reduction in calling rate, the good correlations found between the copy numbers observed in the 1001-gene measurements and those observed in the 140-gene measurements, as well as in

conventional smFISH and bulk RNA sequencing measurements, indicates that the relative abundance of these RNAs can be quantified with the MHD2 encoding scheme.

Simultaneously imaging ~1000 genes in individual cells substantially expanded the ability to detect co-regulated genes. Fig. 10A shows the matrix of pairwise correlation coefficients determined from the cell-to-cell variations in the expression levels of these genes. Using the same hierarchical clustering analysis as described above, ~100 groups of genes with correlated expression were identified. Remarkably, nearly all of these ~100 groups showed statistically significant enrichment of functionally related GO terms (Fig. 10B). These included some of the groups identified in the 140-gene measurements, such as the group associated with cell replication genes and the group associated with cell motility genes (Figs. 10A and 10B, groups 7 and 102), as well as many new groups. The groups identified here included 46 RNA species lacking any previous GO annotations, for which function based on their group association may be hypothesized. For example, KIAA1462 is part of the cell motility group, as also shown in the 140-gene experiments, suggesting a potential role of this gene in cell motility (Fig. 10A, group 102). Likewise, KIAA0355 is part of a new group enriched in genes associated with heart development (Fig. 10A, group 79), and C17orf70 is part of a group associated with ribosomal RNA processing (Fig. 10A, group 22). Using these groupings, cellular functions for 61 transcription factors and other partially annotated proteins of unknown functions may be hypothesized. For example, the transcription factors Z3CH13 and CHD8 are both members of the cell motility group, suggesting their potential role in the transcriptional regulation of cell motility genes.

Fig. 9 shows simultaneous measurements of 1001 genes in single cells using MERFISH with a 14-bit MHD2 code. Fig. 9A shows the localizations of all detected single molecules in a cell colored based on their measured binary words. Inset: the composite, false-colored fluorescent image of the 14 hybridization rounds for the boxed sub-region with numbered circles indicating potential RNA molecules. Circles indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Images of individual hybridization round are shown in Fig. 19A. Fig. 9B is a scatter plot of the average copy number per cell measured in the 1001-gene experiments versus the abundance measured via bulk sequencing. The upper symbols are for the 73% of genes detected with confidence ratios higher than the maximum ratio observed for the misidentification controls. The Pearson correlation coefficient

is 0.76 with a p-value of $3x10^{-133}$. The lower symbols are for the remaining 27% of genes. The Pearson correlation coefficient is 0.65 with a p-value of $3x10^{-33}$. Fig. 9C is a scatter plot of the average copy number for the 107 genes shared in both the 1001-gene measurement with the MHD2 code and the 140-gene measurement with the MHD4 code. The Pearson correlation coefficient is 0.89 with a p-value of $9x10^{-30}$. The dashed line corresponds to the $y = x$ line.

Fig. 10 shows co-variation analysis of the RNA species measured in the 1001-gene measurements. Fig. 10A is a matrix of all pairwise correlation coefficients of the cell-to-cell variation in expression for the measured genes shown with the hierarchical clustering tree. The ~100 identified groups of correlated genes are indicated by shading on the tree. Zoom in of four of the groups described in the text are shown on the right. Fig. 10B is an enrichment of 20 selected, statistically significantly enriched GO terms in the four groups.

Fig. 19 shows decoding and error assessment of the 1001-gene experiments. Fig. 19A shows images of the boxed sub-region of the cell in Fig. 9A for each of 14 hybridization rounds. The final panel is a composite image of these 14 rounds. Circles indicate fluorescent spots that have been identified as potential RNA molecules. Some circles in the composite image indicate unidentifiable molecules, the binary words of which do not match any of the 14-bit MHD2 code words. Fig. 9B shows the corresponding binary word for each of the spots identified in Fig. 9A with the RNA species to which it is decoded. 'unidentified' implies that the measured binary word does not match any of the 1001 code words. Fig. 19C shows the normalized confidence ratios measured for the 985 RNA species (left) and the 16 misidentification control words not targeted to any RNA (right). The normalized confidence ratio is defined as in Fig. 6F. Fig. 19D shows a histogram of the reduction in detected abundance of 107 genes present in both the 1001-gene experiments and the 140-gene experiments. "Fold decrease in copy number" is defined as the average number of RNA molecules per cell for each species measured in the 140-gene experiments divided by the corresponding average number measured in the 1001-gene experiments.

Fig. 18C is a comparison of the average RNA copy numbers per cell measured in the 1001-gene experiments using the MHD2 code to those determined by conventional smFISH for 10 genes. The average ratio of the copy number measured using the MHD2 measurements to that measured using conventional smFISH was 0.30 +/- 0.05 (mean +/- SEM across 10 genes).

The dashed line corresponds to the $y = x$ line and the dotted line corresponds to the $y = 0.30x$ line.

EXAMPLE 9

The above examples illustrate a highly multiplexed detection scheme for systems-level RNA imaging in single cells. Using combinatorial labeling, sequential hybridization and imaging, and two different error-robust encoding schemes, either 140 or 1001 genes in hundreds of individual human fibroblast cells were simultaneously imaged. Of the two encoding schemes presented here, the MHD4 code is capable of both error detection and error correction, and hence can provide a higher calling rate and a lower misidentification rate than the MHD2 code, which instead can only detect but cannot correct errors. MHD2, on the other hand, provides a faster scaling of the degree of multiplexing with the number of bits than MHD4. Other error-robust encoding schemes can also be used for such multiplexed imaging, and experimenters can set the balance between detection accuracy and ease of multiplexing based on the specific requirements of the experiments.

By increasing the number of bits in the code words, it should be possible to further increase the number of detectable RNA species using MERFISH with, for example, a MHD4 or MHD2 code. For example, using the MHD4 code with 32 total bits and four or six '1' bits would increase the number of addressable RNA species to 1,240 or 27,776, respectively. The latter is the approximate scale of the human transcriptome. The predicted misidentification and calling rates are still reasonable for the 32-bit MHD4 code (shown in Figs. 5C and 5D for the MHD4 code with four '1' bits and similar rates were calculated for the MHD4 code with six '1' bits). If more accurate measurements are desired, an additional increase in the number of bits would allow the use of encoding schemes with a Hamming distance greater than 4, further enhancing the error detection and correction capability. While an increase in the number of bits by adding more hybridization rounds would increase the data collection time and potentially lead to sample degradation, these problems could be mitigated by utilizing multiple colors to readout multiple bits in each round of hybridization.

As the degree of multiplexing is increased, it is important to consider the potential increase in the density of RNAs that need to be resolved in each round of imaging. Based on the imaging and sequencing results, it can be estimated that including the whole transcriptome of the IMR90 cells would lead to a total RNA density of ~200 molecules/micrometer^3. Using the

current imaging and analysis methods, 2-3 molecules/ micrometer^3 per hybridization round could be resolved, which would reach a total RNA density of ~20 molecules/micrometer^3 after 32 rounds of hybridization. This density should allow all but the top 10% most expressed genes to be imaged simultaneously or a subset of genes with even higher expression levels to be included. By utilizing more advanced image analysis algorithms to better resolve overlapping images of individual molecules, such as compressed sensing, it is possible to extend the resolvable density by ~4-fold and thus allow all but the top 2% most expressed genes to be imaged all together.

These examples have illustrated the utility of the data derived from highly multiplexed RNA imaging by using co-variation and correlation analysis to reveal distinct sub-cellular distribution patterns of RNAs, to constrain gene regulatory networks, and to predict functions for many previously unannotated or partially annotated genes with unknown functions. Given its ability to quantify RNAs across a wide range of abundances without amplification bias while preserving native context, systems and methods such as MERFISH will allow many applications of *in situ* transcriptomic analyses of individual cells in culture or complex tissues.

EXAMPLE 10

Following are various materials and methods used in the above examples.

Probe design. Each RNA species in the target set was randomly assigned a binary code word either from all 140 possible code words of the 16-bit MHD4 code or from all 1001 possible code words of the 14-bit MHD2 code.

Array-synthesized oligopools were used as templates to make the encoding probes. The template molecule for each encoding probe contained three components: i) a central targeting sequence for *in situ* hybridization to the target RNA, ii) two flanking readout sequences designed to hybridize each of two distinct readout probes, and iii) two flanking primer sequences to allow enzymatic amplification of the probes (Fig. 13). The readout sequences were taken from the 16 possible readout sequences, each corresponding to one hybridization round. The readout sequences were assigned to the encoding probes such that for any RNA species each of the 4 readout sequences were distributed uniformly along the length of the target RNA and appeared at the same frequency. Template molecules for the 140-gene library also included a common 20-nucleotide (nt) priming region between the first PCR primer and the first readout sequence. This priming sequence was used for the reverse transcription step described below.

Multiple experiments were embedded in a single array-synthesized oligopool, and PCR was used to selectively amplify only the oligos required for a specific experiment. Primer sequences for this indexed PCR reaction were generated from a set of orthogonal 25-nt sequences. These sequences were trimmed to 20 nt and selected for i) a narrow melting temperature range (70 °C to 80 °C), ii) the absence of consecutive repeats of 3 or more identical nucleotides, and iii) the presence of a GC clamp, i.e. one of the two 3' terminal bases must be G or C. To further improve specificity, these sequences were then screened against the human transcriptome using BLAST+, and primers with 14 or more contiguous bases of homology were eliminated. Finally, BLAST+ was again used to identify and exclude primers that had an 11-nt homology region at the 3' end of any other primer or a 5-nt homology region at the 3' end of the T7 promoter. The forward primer sequences (Primer 1) were determined as described above, whereas the reverse primers each contain a 20-nt sequence as described above plus a 20-nt T7 promoter sequence to facilitate amplification via *in vitro* transcription (Primer 2). The primer sequences used in the 140-gene and 1001-gene experiments are listed below.

Table 2

Experiment Name	Primer 1 Sequence (Index Primer 1)	Primer 2 Sequence (T7 promoter plus the reverse complement of Index Primer 2)
140-gene Codebook 1	GTTGGTCGGCACTTGGGTGC (SEQ ID NO: 18)	TAATACGACTCACTATAGGGAAAGCCGGTTCATCCG GTGG (SEQ ID NO: 21)
140-gene Codebook 2	CGATGCGCCAATTCCGGTTC (SEQ ID NO: 19)	TAATACGACTCACTATAGGGTGATCATCGCTCGCGG GTTG (SEQ ID NO: 22)
1001-gene	CGCGGGCTATATGCGAACCG (SEQ ID NO: 20)	TAATACGACTCACTATAGGGCGTGGAGGGCATACA ACGC (SEQ ID NO: 23)

30-nt-long readout sequences were created by concatenating fragments of the same orthogonal primer set generated above by combining one 20-nt primer with a 10-nt fragment of another. These readout sequences were then screened, using BLAST+, for orthogonality with the index primer sequences and other readout sequences (no more than 11 nt of homology) and

for potential off-target binding sites in the human genome (no more than 14 nt of homology). Fluorescently labeled readout probes with sequences complementary to the readout sequences were used to probe these readout sequences, one in each hybridization round. All used readout probes sequences are listed below.

Table 3

Bit	Readout probes	
1	CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/	SEQ ID NO: 24
2	CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/	SEQ ID NO: 25
3	ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/	SEQ ID NO: 26
4	CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/	SEQ ID NO: 27
5	GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/	SEQ ID NO: 28
6	AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/	SEQ ID NO: 29
7	CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/	SEQ ID NO: 30
8	ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/	SEQ ID NO: 31
9	CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/	SEQ ID NO: 32
10	GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/	SEQ ID NO: 33
11	CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/	SEQ ID NO: 34
12	GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/	SEQ ID NO: 35
13	TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/	SEQ ID NO: 36
14	CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/	SEQ ID NO: 37
15	GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/	SEQ ID NO: 38
16	GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/	SEQ ID NO: 39

The readout probes used for the 140-gene libraries were probes 1 through 16. The readout probes used for the 1001-gene experiment were probes 1 through 14. /3Cy5Sp/ indicates a 3' Cy5 modification.

To design the central targeting sequences of the encoding probes, the abundance of different transcripts in IMR90 cells using Cufflinks v2.1, total RNA data from the ENCODE project, and human genome annotations from Gencode v18 were complied. Probes were designed from gene models corresponding to the most abundant isoform using OligoArray2.1

with the following constraints: the target sequence region is 30-nt long; the melting temperatures of the hybridized region of the probe and cellular RNA target is greater than 70 °C; there is no cross hybridization targets with melting temperatures greater than 72 °C; there is no predicted internal secondary structures with melting temperatures greater than 76 °C; and there is no contiguous repeats of 6 or more identical nucleotides. Melting temperatures were adjusted to optimize the specificity of these probes and minimize secondary structure while still producing sufficient numbers of probes for the libraries. To decrease computational cost, isoforms were divided into 1-kb regions for probe design. Using BLAST+, all potential probes that mapped to more than one cellular RNA species were rejected. Probes with multiple targets on the same RNA were kept.

For each gene in the 140-gene experiments, 198 putative encoding probe sequences were generated by concatenating the appropriate index primers, readout sequences, and targeting regions as shown in Fig. 13. To address the possibility that concatenation of these sequences introduced new regions of homology to off-target RNAs, BLAST+ was used to screen these putative sequences against all human rRNA and tRNA sequences as well as highly expressed genes (genes with FPKM > 10,000). Probes with greater than 14 nt of homology to rRNAs or tRNAs or greater than 17 nt of homology to highly expressed genes were removed. After these cuts, there were ~192 (with a standard deviation of 2) probes per gene for both MHD4 codebooks used in the 140-gene experiments. The same protocol for the 1001-gene experiments was used, as follows: Starting with 96 putative targeting sequences per gene, ~94 (with a standard deviation of 6) encoding probes per gene were obtained after these additional homology cuts. The number of encoding probes per RNA was decreased for the 1001-gene experiments so that these probes could be synthesized from a single 100,000-member oligopool as opposed to two separate pools. Each encoding probe was designed to contain two of the four readout sequences associated with each code word, hence only half of the bound encoding probes can bind readout probe during any given hybridization round. ~192 or ~94 encoding probes per RNA were used to obtain high signal-to-background ratios for individual RNA molecules. The number of encoding probes per RNA could be substantially reduced but still allow single RNA molecules to be identified. In addition, increasing the number of readout sequences per encoding probe or using optical sectioning methods to reduce the fluorescence background may allow further reduction in the number of the encoding probes per RNA.

Two types of misidentification controls were designed. The first control (blank words) were not represented with encoding probes. The second type of control (no-target words) had encoding probes that were not targeting any cellular RNA. The targeting regions of these probes were composed of random nucleotide sequences subject to the same constraints used to design the RNA targeting sequences described above. Moreover, these random sequences were screened against the human transcriptome to ensure that they contain no significant homology (>14-nt) to any human RNA. The 140-gene measurements contained 5 blank words and 5 no-target words. The 1001-gene measurements contained 11 blank words and 5 no-target words.

Probe synthesis. The encoding probes were synthesized using the following steps, and this synthesis protocol is illustrated in Fig. 13.

Step 1: The template oligopool (CustomArray) was amplified via limited-cycle PCR on a Bio-Rad CFX96 using primer sequences specific to the desired probe set. To facilitate subsequent amplification via *in vitro* transcription, the reverse primer contained the T7 promoter. All primers were synthesized by IDT. This reaction was column purified (Zymo DNA Clean and Concentrator; D4003).

Step 2: The purified PCR products were then further amplified ~200-fold and converted into RNA via a high yield *in vitro* transcription according to the manufacturer's instructions (New England Biolabs, E2040S). Each 20 microliter reaction contained ~1 microgram of template DNA from above, 10 mM of each NTP, 1x reaction buffer, 1x RNase inhibitor (Promega RNasin, N2611) and 2 microliters of the T7 polymerase. This reaction was incubated at 37 °C for 4 hours to maximize yield. This reaction was not purified before the following steps.

Step 3: The RNA products from the above *in vitro* transcription reaction were then converted back into DNA via a reverse transcription reaction. Each 50 microliter reaction contained the unpurified RNA produce from Step 2 supplemented with 1.6 mM of each dNTP, 2 nmol of a reverse transcription primer, 300 units of Maxima H- reverse transcriptase (Thermo Scientific, EP0751), 60 units of RNasin, and a final 1x concentration of the Maxima RT buffer. This reaction was incubated at 50 °C for 45 minutes, and the reverse transcriptase was inactivated at 85 °C for 5 minutes. The templates for the 140-gene libraries contain a common priming region for this reverse transcription step; thus, a single primer was used for this step when creating these probes. Its sequence was CGGGTTTAGCGCCGGAAATG. A common

priming region was not included for the 1001-gene library; thus, the reverse transcription was conducted with the forward primer: CGCGGGCTATATGCGAACCG.

Step 4: To remove the template RNA, 20 microliters of 0.25 M EDTA and 0.5 N NaOH was added to the above reaction to selectively hydrolyze RNA, and the sample was incubated at 95 °C for 10 minutes. This reaction was then immediately purified by column purification using a 100-microgram-capacity column (Zymo Research, D4030) and the Zymo Oligo Clean and Concentrator protocol. The final probes were eluted in 100 microliters of RNase-free deionized water, evaporated in a vacuum concentrator, and then resuspended in 10 μL of encoding hybridization buffer (see below). Probes were stored at -20 °C. Denaturing poly-acrylamide gel electrophoresis and absorption spectroscopy were used to confirm the quality of the probes and revealed that this probe synthesis protocol converts 90-100% of the reverse-transcription primer into full length probe and of the probe that is constructed, 70-80% is recovered during the purification step.

Fluorescently labeled readout probes have sequences complementary to the readout sequences described above and a Cy5 dye attached at the 3' end. These probes were obtained from IDT and HPLC purified.

Sample preparation and labeling with encoding probes. Human primary fibroblasts (American Type Culture Collection, IMR90) were used in this work. These cells are relatively large and flat, facilitating wide-field imaging without the need for optical sectioning. Cells were cultured with Eagle's Minimum Essential Medium. Cells were plated on 22-mm, #1.5 coverslips (Bioptechs, 0420-0323-2) at 350,000 cells/coverslip and incubated at 37 °C with 5% CO_2 for 48-96 hours within petri dishes. Cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1x phosphate buffered saline (PBS; Ambion, AM9625) at room temperature, reduced for 5 minutes with 0.1% w/v sodium borohydride (Sigma, 480886) in water to reduce background fluorescence, washed three times with ice-cold 1x PBS, permeabilized for 2 minutes with 0.5% v/v Triton (Sigma, T8787) in 1x PBS at room temperature, and washed three times with ice cold 1x PBS.

Cells were incubated for 5 minutes in encoding wash buffer comprising 2x saline-sodium citrate buffer (SSC) (Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (NEB, S1402S). 10 microliters of 100 micromolar(140-gene experiments) or 200 micromolar (1001-gene experiments) encoding probes in encoding

hybridization buffer was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18-36 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 1 mg/mL yeast tRNA (Life Technologies, 15401-011) and 10% w/v dextran sulfate (Sigma, D8906-50G).

Cells were then washed with primary encoding wash buffer, incubated at 47 °C for 10 minutes, and this wash was repeated for a total of three times. A 1:1000 dilution of 0.2-micrometer-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8809) in 2xSSC was sonicated for 3 minutes and then incubated with the sample for 5 minutes. The beads were used as fiducial markers to align images obtained from multiple successive rounds of hybridization, as described below. The sample was washed once with 2xSSC, and then post-fixed with 4% v/v paraformaldehyde in 2xSSC at room temperature for 30 minutes. The sample was then washed three times with 2xSSC and either imaged immediately or stored for no longer than 12 hours at 4 °C prior to imaging. All solutions were prepared as RNase-free.

MERFISH imaging with multiple successive rounds of hybridization. The sample coverslip was assembled into a Bioptech's FCS2 flow chamber, and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled 8-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Rainin, Dynamax RP-1). The sample was imaged on a home-built microscope constructed around an Olympus IX-71 body and a 1.45 NA, 100x oil immersion objective and configured for oblique incidence excitation. The objective was heated to 37 °C with a Bioptechs objective heater. Constant focus was maintained throughout the imaging process with a home-built, auto-focusing system. Illumination was provided at 641 nm, 561 nm, and 405 nm using solid state lasers (MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of the Cy5-labeled readout probes, the fiducial beads, and nuclear counterstains, respectively. These lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1) and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a QuadView (Photometrics) using the dichroics T560lpxr, T650lpxr, 750dcxxr (Chroma) and the emission filters ET525/50m, WT59550m-2f, ET700/75m, HQ770lp (Chroma) and imaged with an EMCCD camera (Andor, iXon-897). The camera was configured so that a pixel corresponds to 167 nm in the sample

plane. The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention.

Sequential hybridization, imaging, and bleaching proceeded as follows. 1 mL of 10 nM of the appropriate fluorescently labeled readout probe in readout hybridization buffer (2xSSC; 10% v/v formamide; 10% w/v dextran sulfate, and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 15 minutes. Then 2 mL of readout wash buffer (2xSSC, 20% v/v formamide; and 2 mM vanadyl ribonucleoside complex) was flown across the sample, flow was stopped, and the sample was incubated for 3 minutes. 2 mL of imaging buffer comprising 2xSSC, 50 mM TrisHCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), and 40 microgram/mL catalase (Sigma-Aldrich, C30) was flown across the sample. Flow was then stopped, and then approximately 75 to 100 regions were exposed to ~25 mW 642-nm and 1 mW of 561-nm light and imaged. Each region was 40 micrometers by 40 micrometers. The laser powers were measured at the microscope backport. Because the imaging buffer is sensitive to oxygen, the ~50 mL of imaging buffer used for a single experiment was made fresh at the beginning of the experiment and then stored under a layer of mineral oil throughout the measurement. Buffer stored in this fashion was stable for more than 24 hours.

After imaging, the fluorescence of the readout probes was extinguished via photobleaching. The sample was washed with 2 mL of photobleaching buffer (2xSSC and 2 mM vanadyl ribonucleoside complex), and each imaged region of the sample was exposed to 200 mW of 641-nm light for 3 s. To confirm the efficacy of this photobleaching treatment, imaging buffer was reintroduced, and the sample was imaged as described above.

The above hybridization, imaging, and photobleaching process was repeated either 16 times for the 140-gene measurements using the MHD4 code or 14 times for the 1001-gene measurements using the MHD2 code. An entire experiment was typically completed in ~20 hours.

Following completion of imaging, 2 mL of a 1:1000 dilution of Hoescht (ENZ-52401) in 2xSSC was flown through the chamber to label the nuclei of the cells. The sample was then washed immediately with 2 mL of 2xSSC followed by 2 mL of imaging buffer. Each region of the sample was then imaged once again with ~1 mW of 405-nm light.

Because cells were imaged using wide-field imaging with oblique-incidence illumination, without optical sectioning and z-scanning, the fraction of individual RNA species that was outside the axial range of the imaging geometry was quantified for 6 different RNA species using conventional smFISH. For this purpose, these cells were optically sectioned by collecting stacks of images at different focal depths through the entire depth of the cells. The images were aligned in consecutive focal planes and then computed for each cell the fraction of RNAs that were detected in the three-dimensional stack but not in the basal focal plane. It was found that only a small fraction, 15% +/- 1% (Mean +/- SEM across six different RNA species) of RNA molecules were outside the imaging range of a fixed focal plane without z-scanning. These measurements also confirmed that the excitation geometry illuminated the full depth of the cells. Any optical sectioning technique could be employed in MERFISH to allow the imaging of RNAs in thicker cells or tissues.

Construction of measured words. Fluorescent spots were identified and localized in each image using a multi-Gaussian-fitting algorithm assuming a Gaussian with a uniform width of 167 nm. This algorithm was used to allow partially overlapping spots to be distinguished and individually fit. RNA spots were distinguished from background signal, i.e. signal arising from probes bound non-specifically, by setting the intensity threshold required to fit a spot with this software. Due to variation in the brightness of spots between rounds of hybridization, this threshold was adjusted appropriately for each hybridization round to minimize the combined average of the 1-->0 and 0-->1 error rates across all hybridization rounds (140-gene measurements) or to maximize the ratio of the number of measured words with four '1' bits to those with three or five '1' bits (1001-gene measurements). The location of the fiducial beads was identified in each frame using a faster single-Gaussian fitting algorithm.

Images of the same sample region in different rounds of hybridization were registered by rotating and translating the image to align the two fiducial beads within the same image that were most similar in location after a coarse initial alignment via image correlation. All images were aligned to a coordinate system established by the images collected in the first round of hybridization. The quality of this alignment was determined from the residual distance between five additional fiducial beads, and alignment error was typically ~20 nm.

Fluorescence spots in different hybridization rounds were connected into a single string, corresponding to a potential RNA molecule, if the distance between spots was smaller than 1

pixel (167 nm). For each string of spots, the on-off sequence of fluorescent signals in all hybridization rounds were used to assign a binary word to the potential RNA molecule, in which '1' was assigned to the hybridization rounds that contained a fluorescent signal above threshold and '0' was assigned to the other hybridization rounds. Measured words were then decoded into RNA species using the 16-bit MHD4 code or the 14-bit MHD2 code discussed above. In the case of the 16-bit MHD4 code, if the measured binary word matched the code word of a specific RNA perfectly or differed from the code word by one single bit, it was assigned to that RNA. In the case of the 14-bit MHD2 code, only if the measured binary word matched the code word of a specific RNA perfectly, was it assigned to that RNA. To determine the copy number per cell, the number of each RNA species was counted in individual cells within each 40 micrometer by 40 micrometer imaging area. It is noted that this number accounts for the majority but not all RNA molecules within a cell because a fraction of the cell could be outside the imaging area or focal depth. Tiling images of adjacent areas and adjacent focal planes could be employed to improve the counting accuracy.

In the 140-gene experiments, some regions of the cell nucleus occasionally contained too much fluorescence signal to properly identify individual RNA spots. In the 1001-gene experiments, the cell nucleus generally contained too much fluorescent signal to allow identification of individual RNA molecules. These bright regions were excluded from all subsequent analysis. This work focuses on mRNAs, which are enriched in the cytoplasm. To estimate the fraction of mRNAs missed by excluding the nucleus region, conventional smFISH was used to quantify the fraction of molecules found inside the nucleus for six different mRNAs species. It was found that only 5% +/- 2% (Mean +/- SEM across six RNA species) of these RNA molecules are found in the nucleus. Employment of super-resolution imaging and/or optical sectioning could potentially allow individual molecules in these dense nucleus regions to be identified, which would be particularly useful for probing those non-coding RNAs that are enriched in the nucleus.

smFISH measurements of individual genes. Pools of 48 fluorescently-labeled (Quasar 670) oligonucleotide probes per RNA were purchased from Biosearch Technologies. 30-nt probe sequences were taken directly from a random subset of the targeting regions used for the multiplexed measurements. Cells were fixed and permeabilized as described above. 10 microliters of 250 nM oligonucleotide probes in encoding hybridization buffer (described above)

was added to the cell-containing coverslip and spread uniformly by placing another coverslip on top of the sample. Samples were then incubated in a humid chamber inside a 37 °C-hybridization oven for 18 hours. Cells were then washed with encoding wash buffer (described above) at 37 °C for 10 minutes, and this wash was repeated for a total of three times. The sample was then washed three times with 2xSSC and imaged in imaging buffer using the same imaging geometry as described above for MERFISH.

Bulk RNA sequencing. Total RNA was extracted from IMR90 cells cultured as above using the Zymo Quick RNA MiniPrep kit (R1054) according to the manufacturer's instructions. polyA RNA was then selected (NEB; E7490), and a sequencing library was constructed using the NEBNext Ultra RNA library preparation kit (NEB; E7530), amplified with custom oligonucleotides, and 150-bp reads were obtained from on a MiSeq. These sequences were aligned to the human genome (Gencode v18) and isoform abundance was computed with cufflinks.

Calculation of the predicted scaling and error properties of different encoding schemes. Analytic expressions were derived for the dependence of the number of possible code words, the calling rate, and the misidentification rate on *N*. The calling rate is defined as the fraction of RNA molecules that are properly identified. The misidentification rate is defined as the fraction of RNA molecules that are misidentified as a wrong RNA species. For encoding schemes with an error-detection capability, the calling rate and misidentification rate does not add up to 1 because a fraction of the molecules not called properly can be detected as errors and discarded and, hence, not misidentified as a wrong species. These calculations assume that the probability of misreading bits is constant for all hybridization rounds but differs for the 1-->0 and 0-->1 errors. Experimentally measured average 1-->0 and 0-->1 error rates (10% and 4% respectively) were used for the estimates shown in Figs. 5B-5D. For simplicity, the word corresponding to all '0's was not removed from calculations.

For the simple binary encoding scheme in which all possible *N*-bit binary words are assigned to unique RNA species, the number of possible code words is 2^N. The number of words that could be used to encode RNA is actually $2^N - 1$ because the code word '00...0' does not contain detectable fluorescence in any hybridization round, but for simplicity the word corresponding to all '0's was not removed from subsequent calculations. The error introduced

by this approximation is negligible. For any given word with m '1's and N–m '0's the probability of measuring that word without error (the fraction of RNAs that is properly called) is:

$$(1-p_1)^m(1-p_0)^{N-m}, \quad (1)$$

where p_1 is 1-->0 error rate and p_0 is 0-->1 error rate per bit. Because different words in this simple binary encoding scheme can have different numbers of '1' bits, the calling rate for different words will differ if $p_1 \neq p_0$. The average calling rate, reported in Fig. 5C, was determined from the weighted average of the value of Eq. (1) for all words. This weighted average is:

$$\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}, \quad (2)$$

where $\binom{N}{m}$ is the binomial coefficient and corresponds to the number of words with m '1' bits in this encoding scheme. Since in this encoding scheme every error produces a binary word that encodes a different RNA, the average misidentification rate for this encoding scheme, reported in Fig. 5D, follows directly from (2):

$$1-\frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}(1-p_1)^m(1-p_0)^{N-m}. \quad (3)$$

To calculate the scaling and error properties of the extended Hamming distance 4 (HD4) code, the generator matrix for the desired number of data bits using standard methods was first created. The generator matrix determines the specific words that are present in a given encoding scheme and was used to directly determine the number of encoded words as a function of the number of bits. In this encoding scheme, the calling rate corresponds to the fraction of words measured without error as well as the fraction of words measured with a single-bit error. For code words with m '1' bits, this fraction is determined by the following expression:

$$(1-p_1)^m(1-p_0)^{N-m} + m\,p_1{}^1(1-p_1)^{m-1}(1-p_0)^{N-m} + (N-m)\,p_0{}^1(1-p_1)^m(1-p_0)^{N-m-1} \quad (4)$$

where the first term is the probability of not making any errors, the second term corresponds to the total probability of making one 1-->0 error at any of the m '1' bits without making any other 0-->1 errors, and the final term corresponds to the total probability of making one 0-->1 error at any of the N-m '0' bits without making any 1-->0 errors. Because the number of '1' bits can differ between words in this encoding scheme, the average calling rate reported in Fig. 5C was computed from a weighted average over Eq. (4) for different values of m. The weight for each

term was determined from the number of words that contain m '1' bits as determined from the generator matrix described above.

Because RNA-encoding words are separated by a minimum Hamming distance of 4, at least 4 errors are required to switch one word into another. If error correction is applied, then 3 or 5 errors could also convert one RNA into another. Thus, the misidentification rate from all possible combinations of 3-bit, 4-bit and 5-bit errors was estimated for code words with m '1' bits. Technically, >5-bit errors could also convert one RNA into another, but the probability of making such errors is negligible because of the small per-bit error rate. This expression was approximated with:

$$\begin{aligned}
&\sum_{i=0}^{4}\binom{m}{i}\binom{N-m}{4-i}p_1^i p_0^{4-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(4-i)} + \\
&\sum_{i=0}^{3}\binom{m}{i}\binom{N-m}{3-i}p_1^i p_0^{3-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(3-i)} + \\
&\sum_{i=0}^{5}\binom{m}{i}\binom{N-m}{5-i}p_1^i p_0^{5-i}(1-p_1)^{m-i}(1-p_0)^{N-m-(5-i)} +
\end{aligned} \tag{5}$$

The first sum corresponds to all of the ways in which exactly four mistakes can be made. Similarly, the second and third sums correspond to all of the ways in which exactly three or five mistakes can be made. Eq. (5) provides an upper bound for the misidentification rate because not all three, four, or five bit errors produce a word that matches or would be corrected to another legitimate word. Again because the number of '1' bits can differ between words, the average misidentification rate reported in Fig. 5D is calculated as a weighted average of Eq. (5) over the number of words that have m '1' bits.

To generate the MHD4 code where the number of '1' bits for each code word is set to 4, the HD4 codes were first generated as described above, and then all code words that did not contain four '1's were removed. The calling rate of this code, reported in Fig. 5C, was directly calculated from Eq. (4) but with m = 4 because all code words in this code have four '1' bits. The misidentification rate of this code, reported in Fig. 5D, was calculated by modifying Eq. (5) with the following considerations: (i) the number of '1' bits, m, was set to 4 and (ii) errors that produce words that do not contain three, four, or five '1' bits were excluded. Thus, the expression in Eq. (5) was simplified to

$$\begin{aligned}
&\binom{4}{2}\binom{N-4}{2}p_1^2p_0^2(1-p_1)^2(1-p_0)^{N-6}+\\
&\binom{4}{1}\binom{N-4}{2}p_1^1p_0^2(1-p_1)^3(1-p_0)^{N-6}+\binom{4}{2}\binom{N-4}{1}p_1^2p_0^1(1-p_1)^2(1-p_0)^{N-5}\\
&\binom{4}{2}\binom{N-4}{3}p_1^2p_0^3(1-p_1)^2(1-p_0)^{N-7}+\binom{4}{3}\binom{N-4}{2}p_1^3p_0^2(1-p_1)^1(1-p_0)^{N-6}
\end{aligned} \tag{6}$$

Again, this expression is an upper bound on the actual misidentification rate because not all words with four '1's are valid code words.

Estimates of the 1-->0 and 0-->1 error rates for each hybridization round. To compute the probability of misreading a bit at a given hybridization round, the error correcting properties of the MHD4 code were used. Briefly, the probabilities of 1-->0 or 0-->1 errors were derived in the following way. Let the probability of making an error at the *i*th bit, i.e. *i*th hybridization round, be p_i and the actual number of RNA molecules of the given species be *A*, then the number of exact matches for this RNA will be $W_E = A\prod_{i=1}^{16}(1-p_i)$ and the number of one-bit error corrected matches for this RNA corresponding to errors at the *i*th bit will be $W_i = A\frac{p_i}{(1-p_i)}\prod_{j=1}^{16}(1-p_j)$. The p_i can be directly derived from the ratio: $W_i/W_E = \frac{p_i}{(1-p_i)}$. This ratio assumes that the one-bit error-corrected counts were only generated from single-bit errors from the correct word and that multi-error contamination from other RNA words is negligible. Given that the error rate per hybridization round is small and that it takes at least three errors to convert one RNA-encoding word into a word that would be misidentified as another RNA, the above approximation should be a good one.

To compute the average 1-->0 or 0-->1 error probabilities for each of the 16 hybridization rounds, the above approach was used to calculate the per-bit error rates for each bit of every gene, and these errors were sorted based on whether they correspond to a 1-->0 or a 0-->1 error, and the average of these errors for each bit weighted by the number of counts observed for the corresponding gene was taken.

Estimates of the calling rate for individual RNA species from actual imaging data. With the estimates of the 1-->0 or 0-->1 error probabilities for each round of hybridization as determined above, it is possible to estimate the calling rate for each RNA based on the specific

word used to encoded it. Specifically, the fraction of an RNA species that is called correctly is determined by

$$\prod_{i=1}^{N}(1-p_i)+\sum_{j=1}^{N}\frac{p_j}{(1-p_j)}\prod_{i=1}^{N}(1-p_i), \qquad (7)$$

where the first term represent the probability of observing an exact match of the code word and the second term represent the probability of observing an error-corrected match (i.e. with one-bit error). The values of the per-bit error rate p_i for each RNA species are determined by the specific code word for that RNA and the measured 1-->0 or 0-->1 error rates for each round of hybridization. If the code word of the RNA contains a '1' in the *i*th bit, then p_i is determined from the 1-->0 error rate for the *i*th hybridization round; if the word contains a '0' in the *i*th bit, p_i is determined from the 0-->1 error rate for the *i*th hybridization round.

Hierarchical clustering analysis of the co-variation in RNA abundance. Hierarchical clustering of the co-variation in gene expression for both the 140-gene and 1001-gene experiments was conducted as follows. First, the distance between every pair of genes was determined as 1 minus the Pearson correlation coefficient of the cell-to-cell variation of the measured copy numbers of these two RNA species, both normalized by the total RNA counted in the cell. Thus, highly correlated genes are 'closer' to one another and highly anti-correlated genes are 'further' apart. An agglomerative hierarchical cluster tree was then constructed from these distances using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Specifically, starting with individual genes, hierarchical clusters were constructed by identifying the two clusters (or individual genes) that are closest to one another according to the arithmetic mean of the distances between all inter-cluster gene pairs. The pairs of clusters (or individual genes) with the smallest distance are then grouped together and the process is repeated. The matrix of pairwise correlations was then sorted based on the order of the genes within these trees.

Groups of genes with substantial co-variations were identified by selecting a threshold on the hierarchical cluster tree (indicated by the dashed lines in Figs. 7D and 10A) that produced approximately 10 groups of genes each of which contains at least 4 members for the 140-gene experiments or approximately 100 groups each of which contains at least 3 members for the 1001-gene experiments. It is noted that one can change the threshold in order to identify either more tightly coupled smaller groups or larger groups with relatively loose coupling.

A probability value for the confidence that a gene belongs to a specific group was determined by computing the difference between the average correlation coefficient between that gene and all other members of that group and the average correlation coefficient between that gene and all other measured genes outside that group. The significance (p-value) of this difference was determined with the student's t-test.

Because hierarchical clustering is inherently a one-dimensional analysis, i.e. any given genes can only be a member of a single group, this analysis does not allow all correlated gene groups to be identified. Higher dimension analysis, such as principal component analysis or k-means clustering, could be used to identify more co-varying gene clusters.

Analysis of RNA spatial distributions. To identify genes that have similar spatial distributions, each of the measured cells was subdivided into 2x2 regions and calculated the fraction of each RNA species present in each of these bins. To control for the fact that some regions of the cell naturally contain more RNA than others, the enrichment was calculated for each gene , i.e., the ratio of the observed fraction in a given region for a given RNA species to the average fraction observed for all genes in that same region. For each pair of RNA species, the Pearson correlation coefficient of the region-to-region variation in enrichment of these two RNA species for each cell was determined and the correlation coefficients were averaged over ~400 cells imaged in 7 independent data sets. RNA species were then clustered based on these average correlation coefficients using the same hierarchical clustering algorithm described above. Because of the large number of cells used for the analysis, it was found that the coarse spatial binning (2x2 regions per cell) was sufficient to capture the spatial correlation between genes and finer binning did not produce more significantly correlated groups.

To measure the distances of genes from the nuclei and from the cell edge, brightness thresholds on the cell images were first used to segment the nuclei and the cell edges identified. The distance from every RNA molecule to the nearest part of the nucleus and nearest part of the cell edge was then determined. For each data set, the average distance for each RNA species averaged over all the cells measured was then computed. These distances were averaged for the group I genes, group II genes or all genes. Only those RNA species with at least 10 counts per cell were used in this analysis to minimize statistical error on the distance values.

Gene ontology (GO) analysis. Groups of genes were selected from the hierarchical trees as discussed above. A collection of GO terms was determined for all measured RNA species as

well as the RNA species associated with each group from the most recent human GO annotations using both the annotated GO terms and terms immediately upstream or downstream of the found annotations. The enrichment of these annotations was calculated from the ratio of the fraction of genes within each group that have this term to the fraction of all measured genes that have this term and the p-value for this enrichment was calculated via the hypergeometric function. Only statistically significantly enriched GO terms with a p-value less than 0.05 were considered.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in

one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

When the word "about" is used herein in reference to a number, it should be understood that still another embodiment of the invention includes that number not modified by the presence of the word "about."

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

ABSTRACT

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Encoding Scheme

RNA target	codeword	probe sequence
mRNA I	110	
mRNA II	011	
mRNA III	101	

Primary sites (RNA tag)
Secondary site 1
Secondary site 2
Secondary site 3

Fig. 1A

b) **Code-Reading Scheme 1: Primaries and Secondaries**

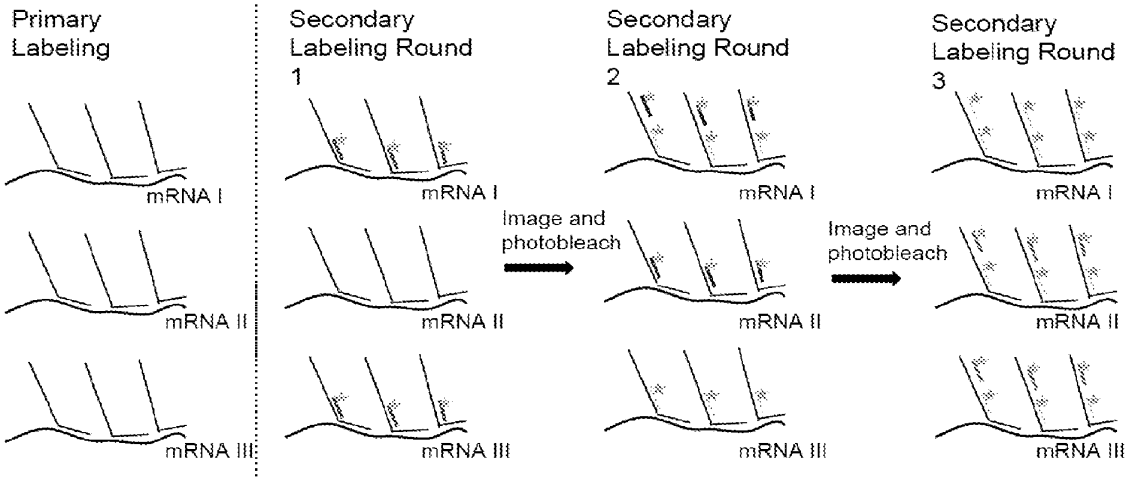


Fig. 1B

Code-Reading Scheme 2: Primaries Only

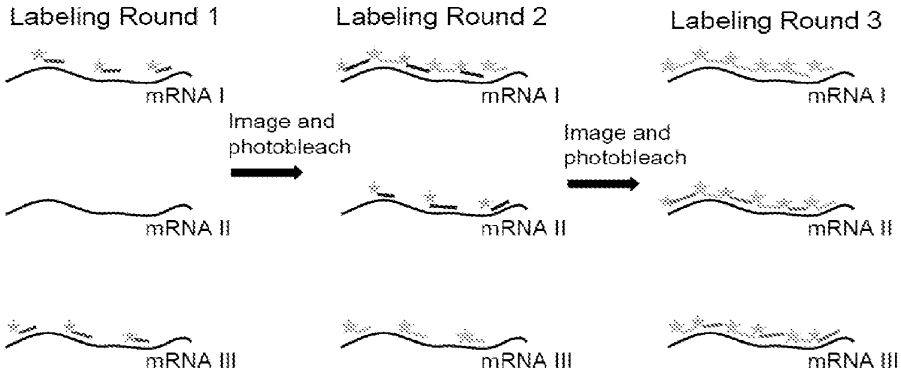


Fig. 1C

Fig. 2A

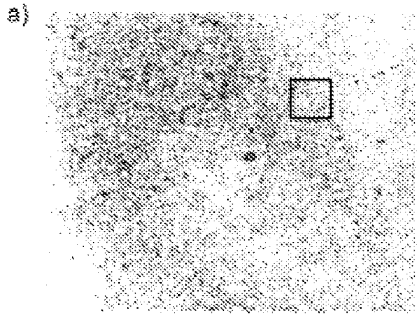


Fig. 2B

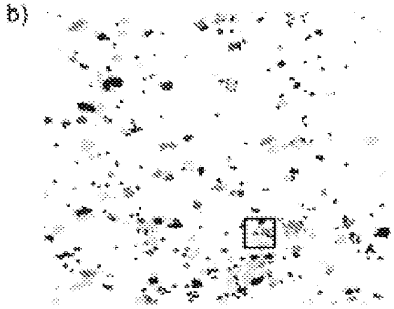
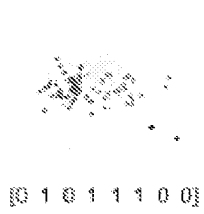


Fig. 2C



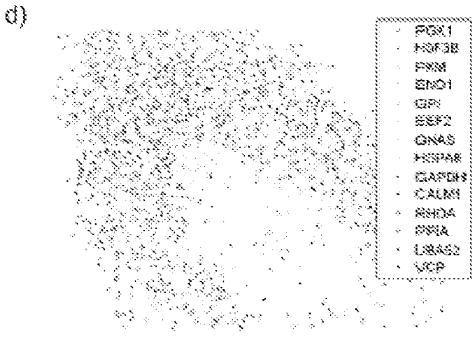


Fig.2D

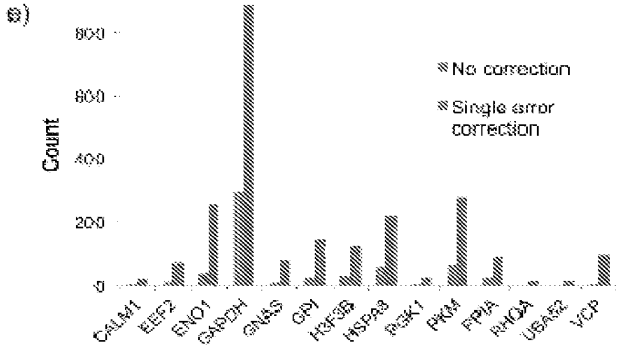


Fig. 2E

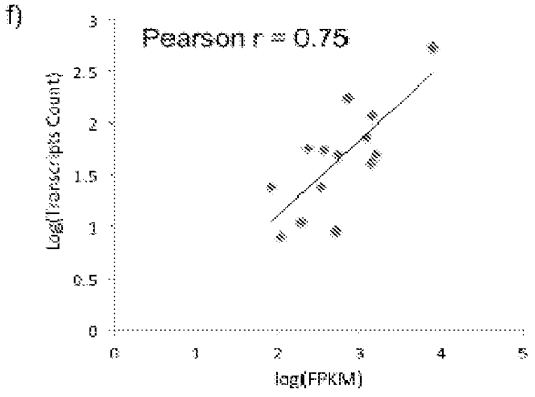


Fig. 2F

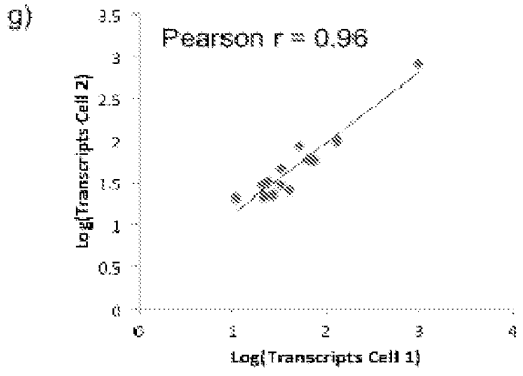


Fig. 2G

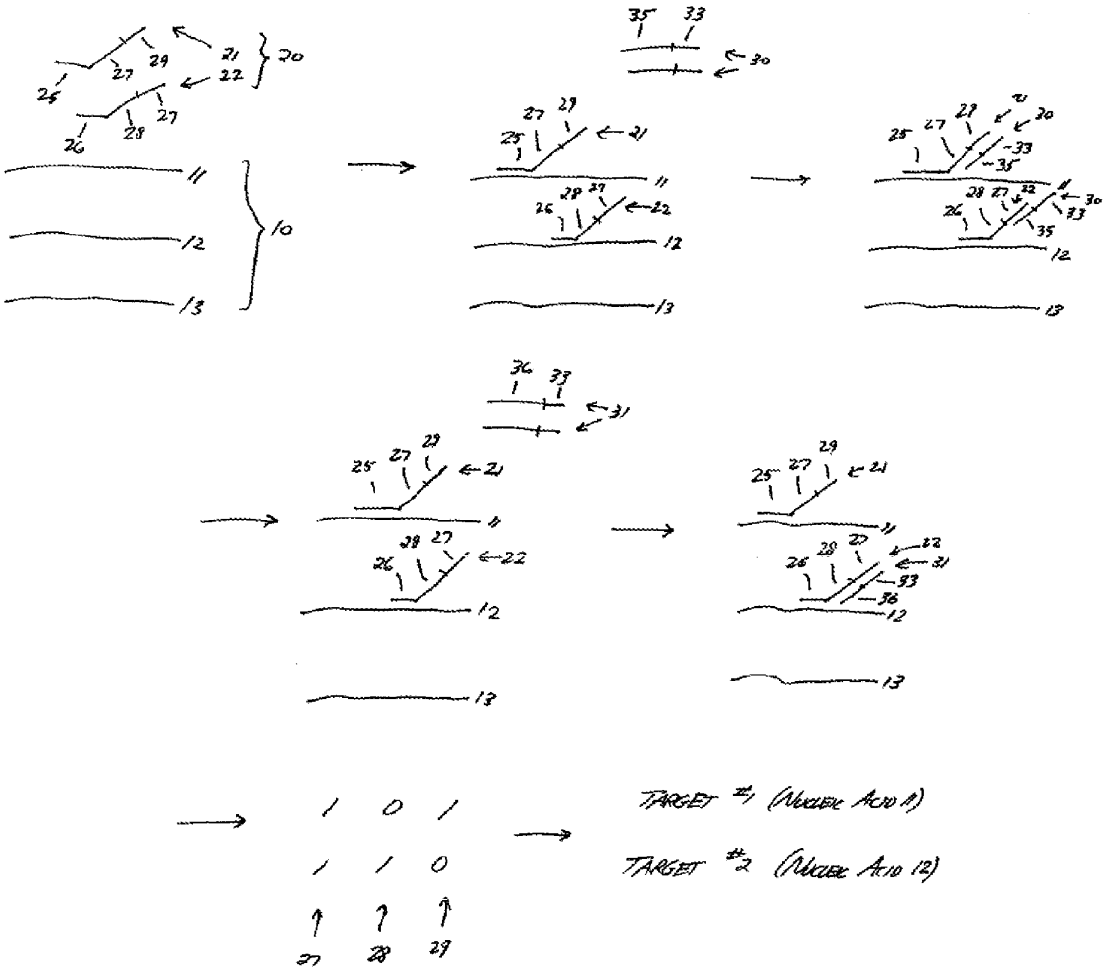

10
11
12
13
20
21
22
25
26
27
28
29
30
31
33
35
36
1 0 1
1 1 0
27 28 29
TARGET #1 (NUCLEIC ACID 1)
TARGET #2 (NUCLEIC ACID 12)

Fig. 3A

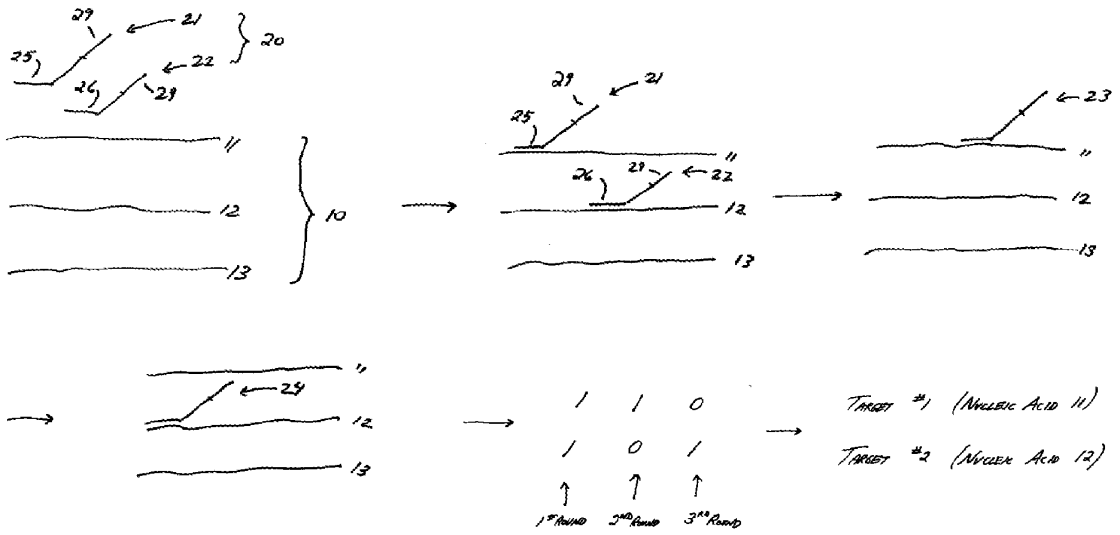


Fig. 3B

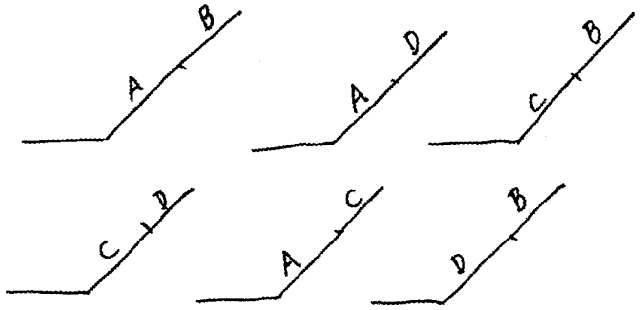


Fig. 4A

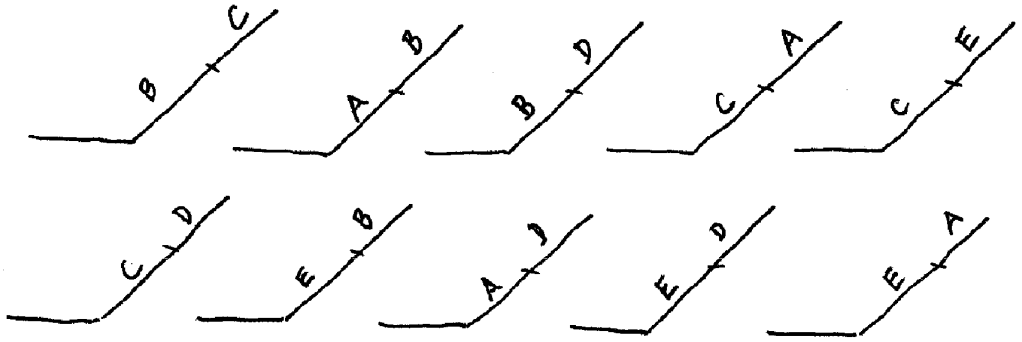


Fig. 4B

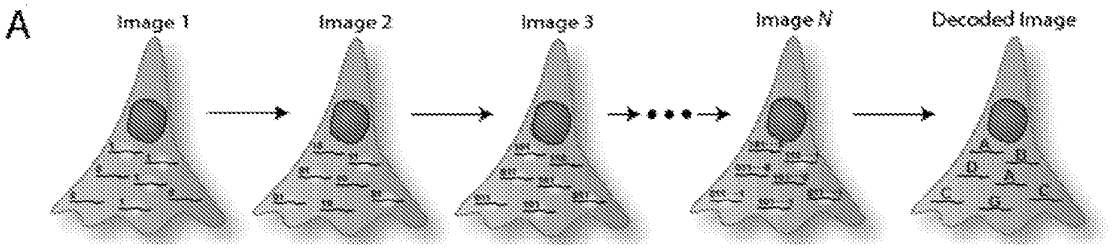


Fig. 5A

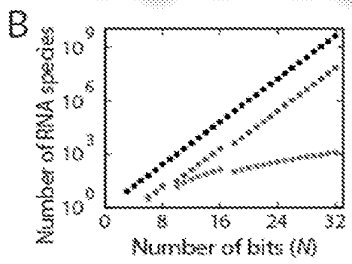


Fig. 5B

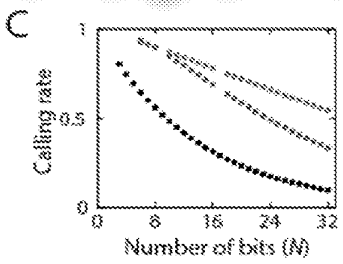


Fig. 5C

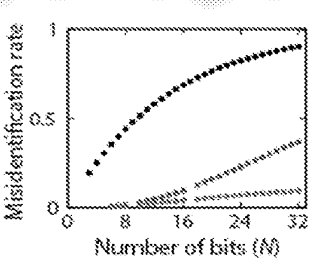


Fig. 5D

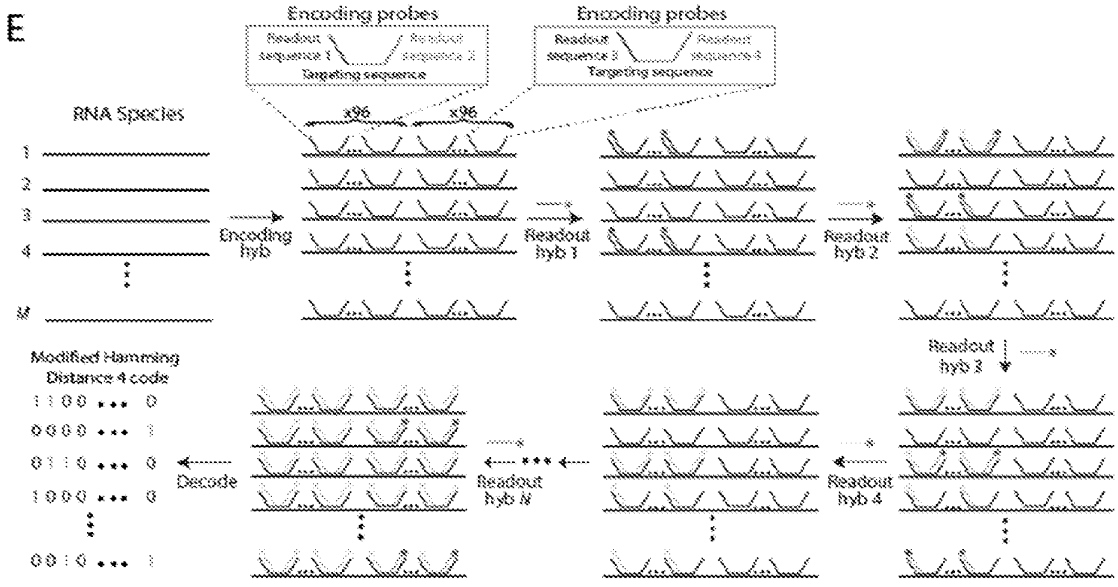


Fig. 5E

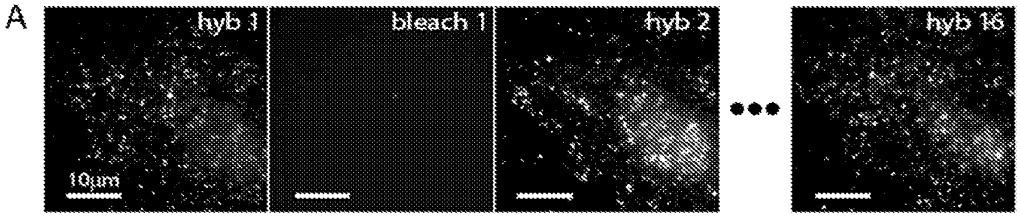


Fig. 6A

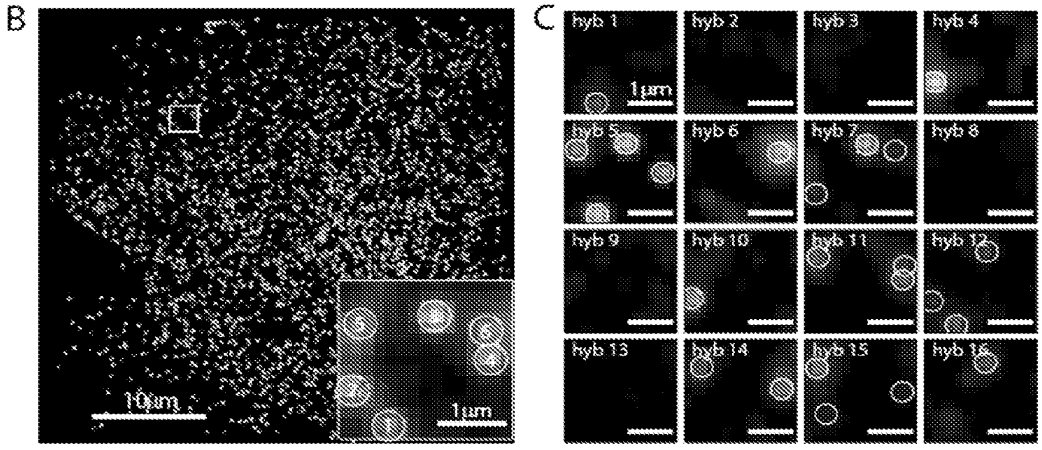


Fig. 6B

Fig. 6C

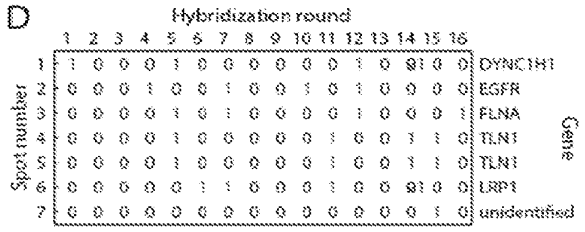


Fig. 6D

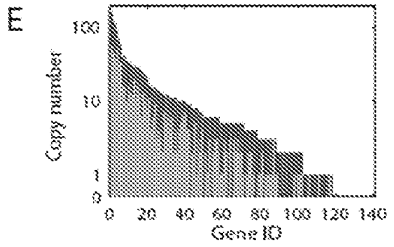


Fig. 6E

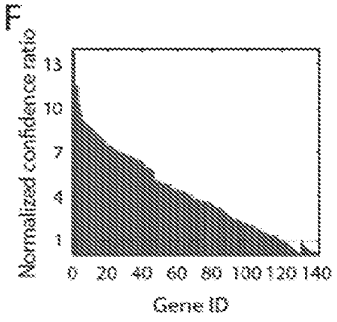


Fig. 6F

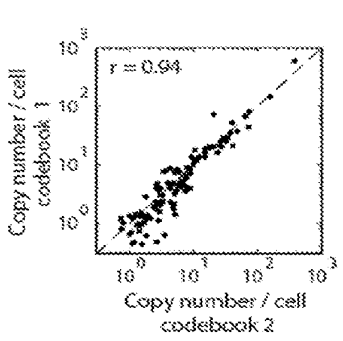


Fig. 6G

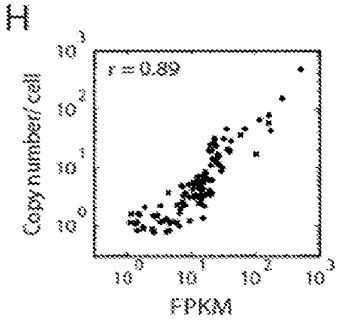


Fig. 6H

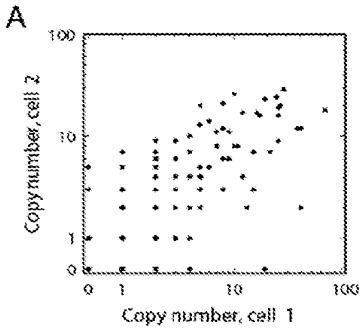


Fig. 7A

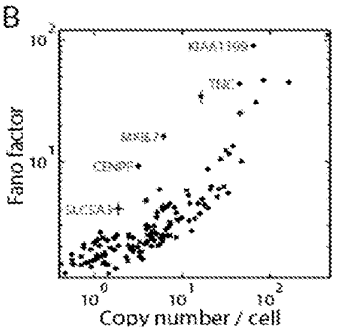


Fig. 7B

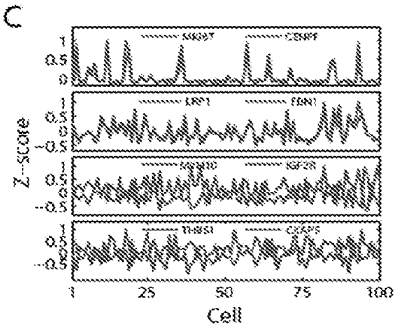


Fig. 7C

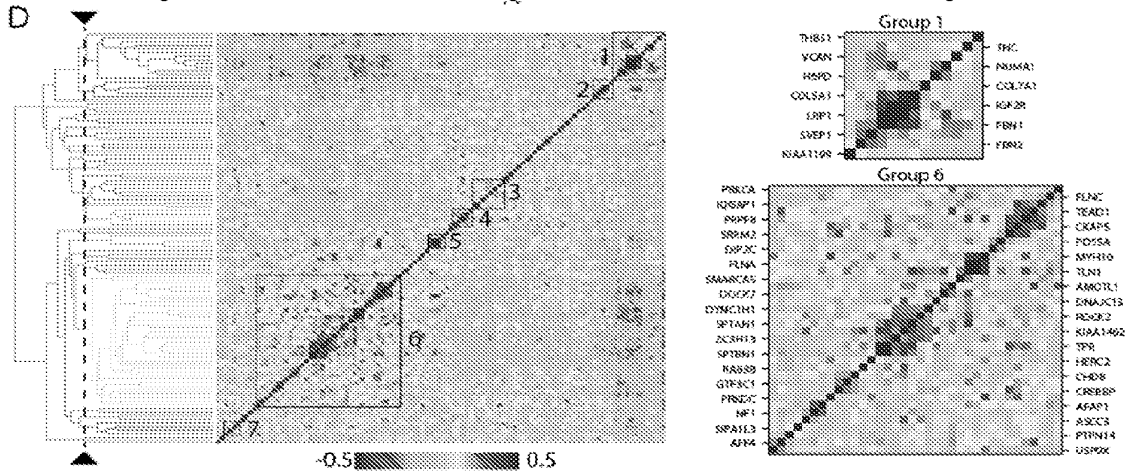


Fig. 7D

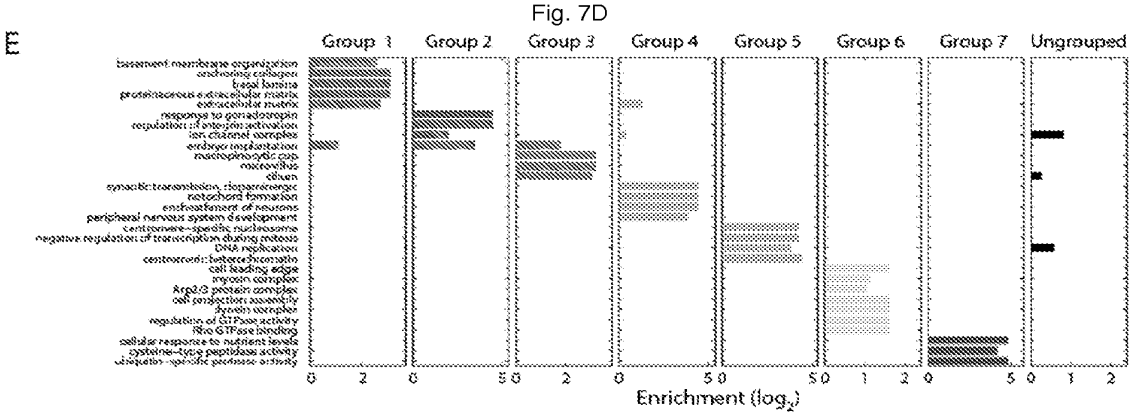


Fig. 7E

A

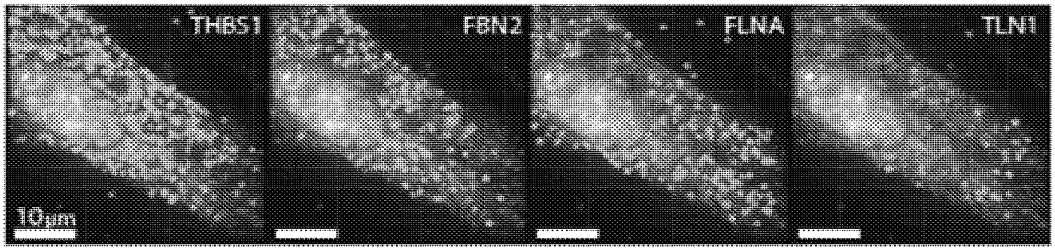


Fig. 8A

B

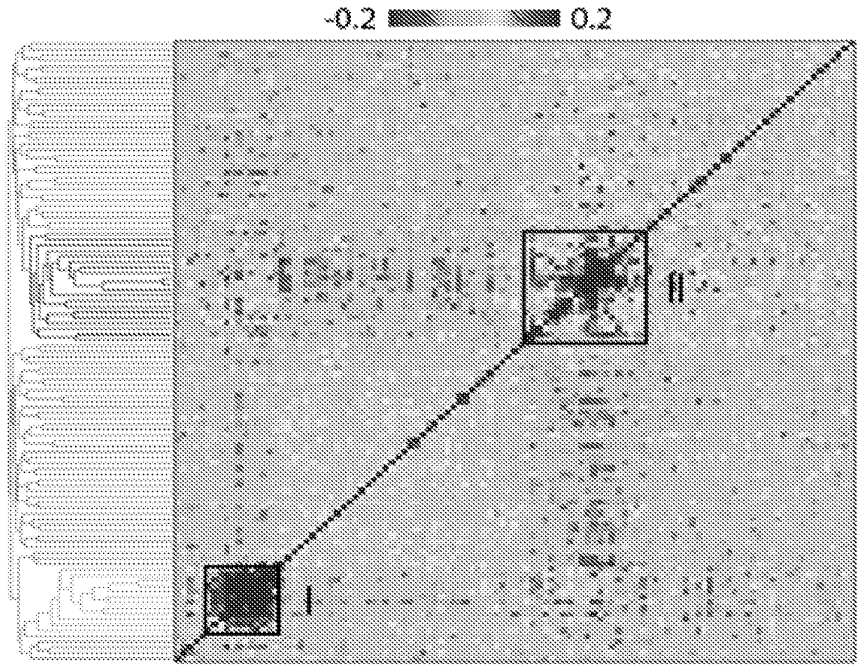


Fig. 8B

C

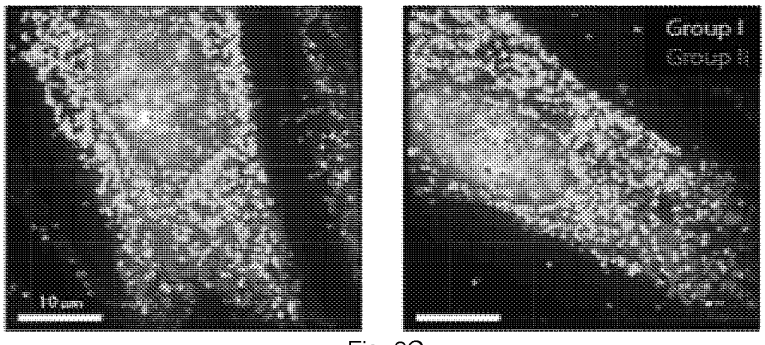


Fig. 8C

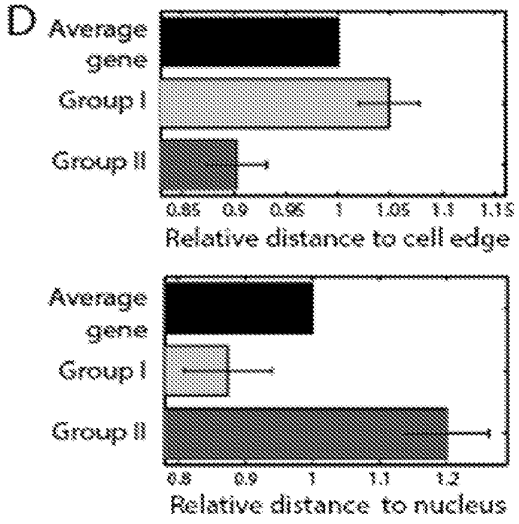


Fig. 8D

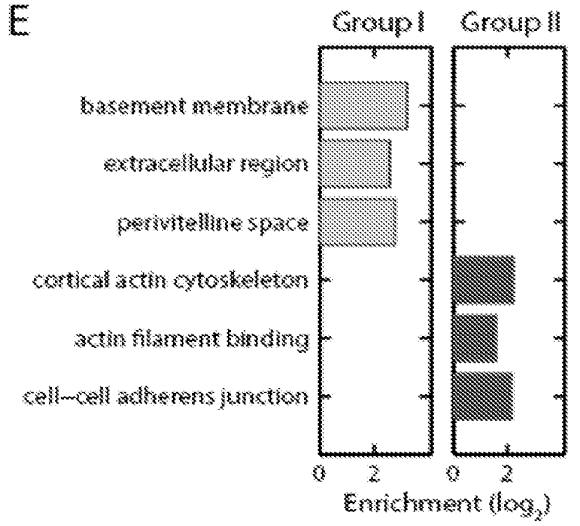


Fig. 8E

Fig. 9A

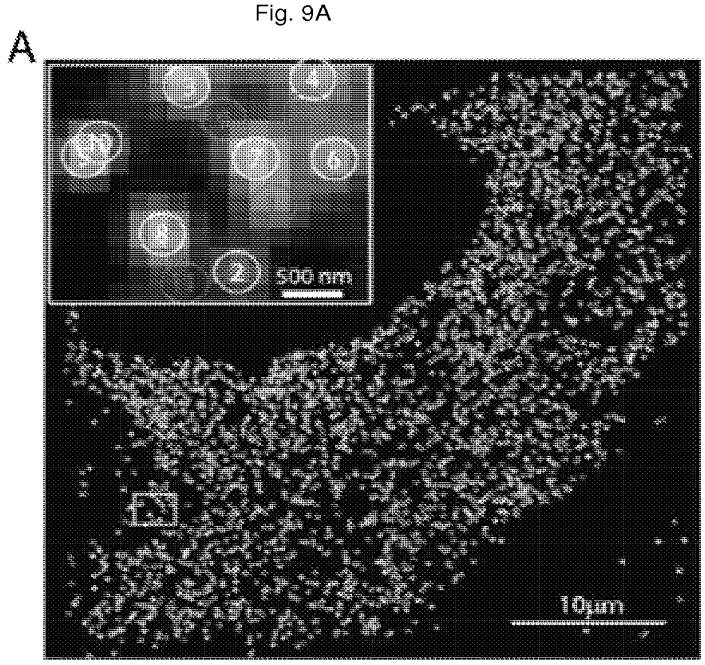


Fig. 9B

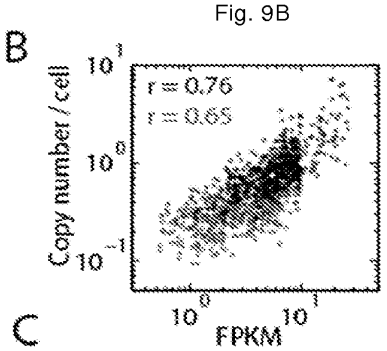
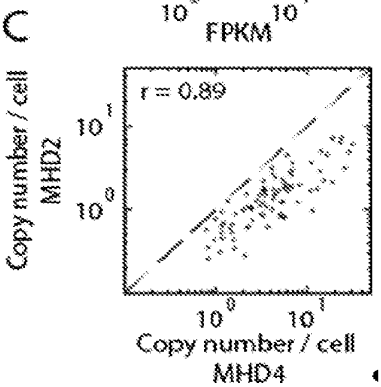


Fig. 9C



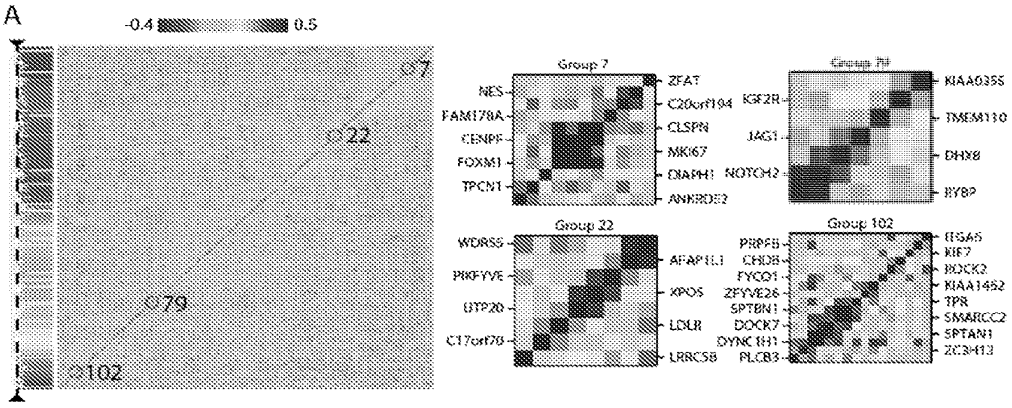


Fig. 10A

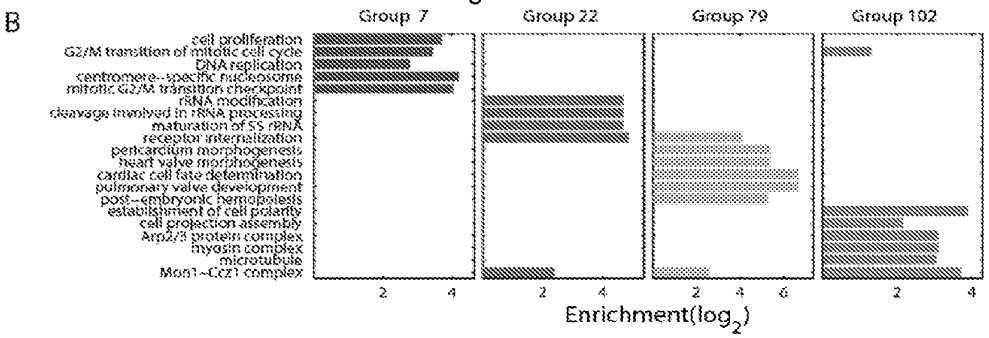


Fig. 10B

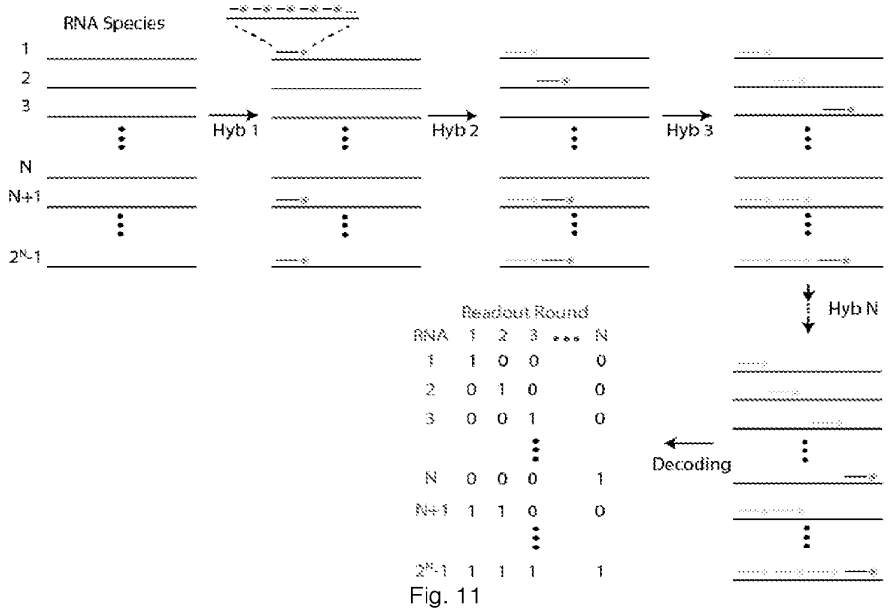


Fig. 11

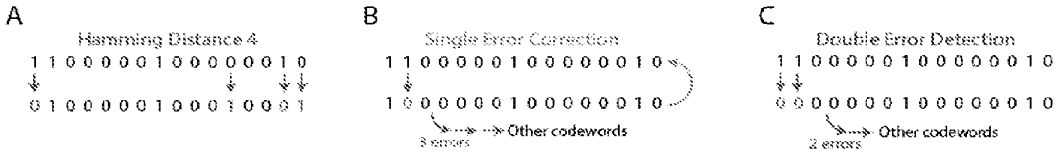


Fig. 12

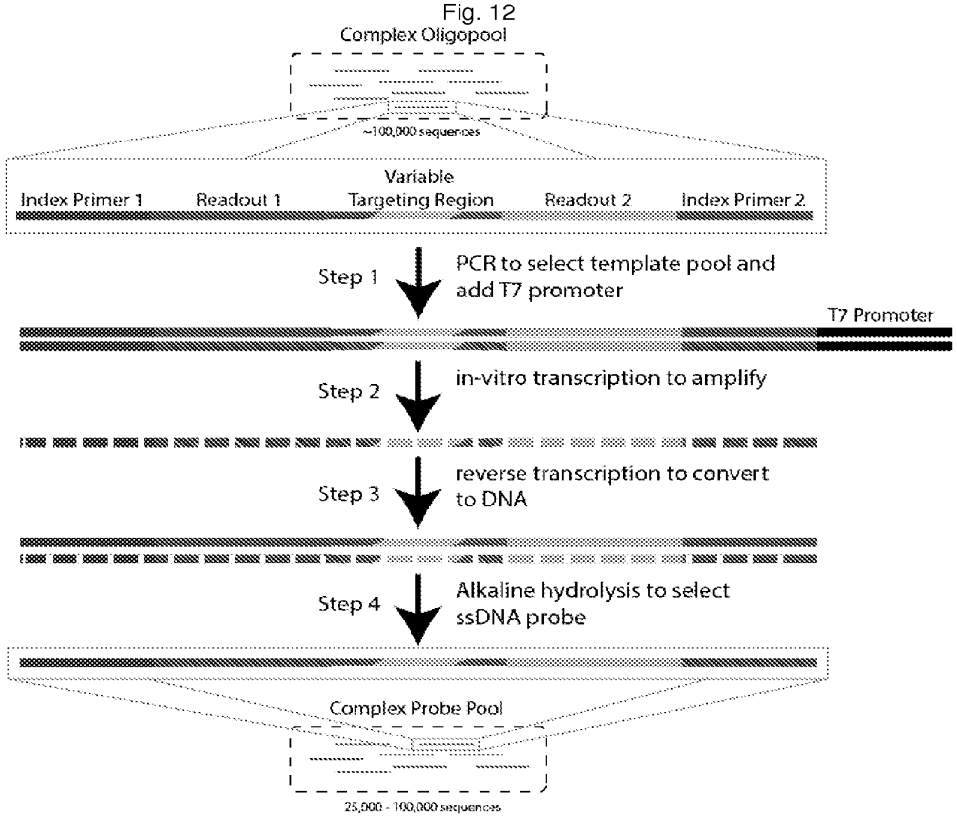


Fig. 13

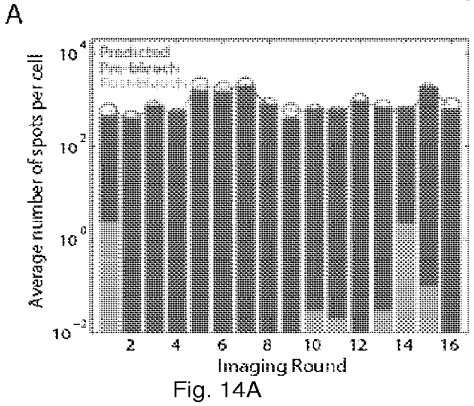


Fig. 14A

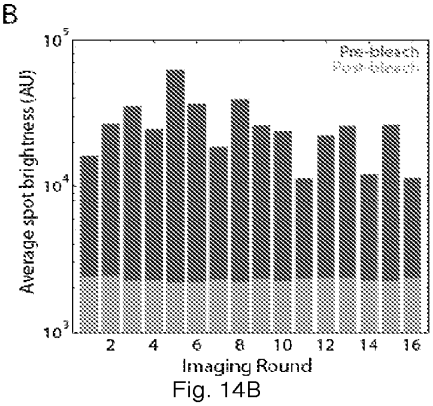


Fig. 14B

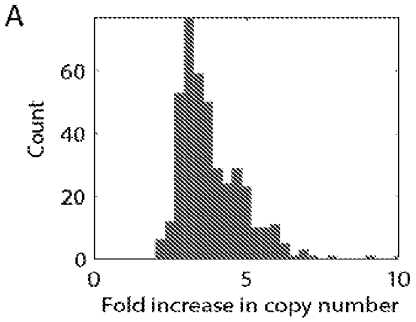


Fig. 15A

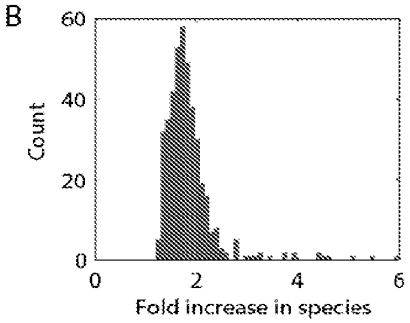


Fig. 15B

Codebook 1

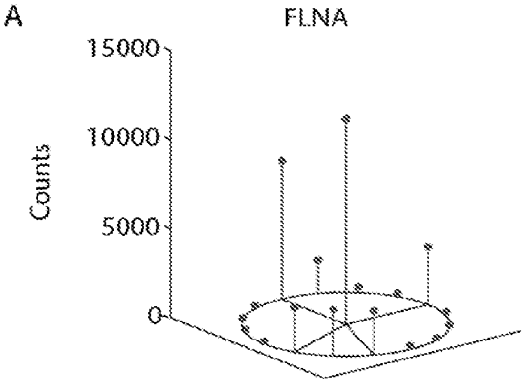


Fig. 16A

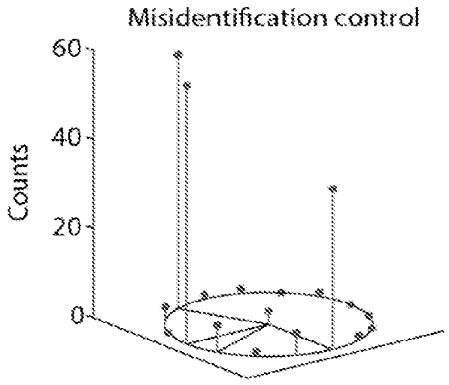


Fig. 16B

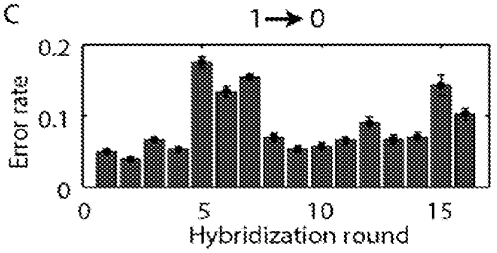


Fig. 16C

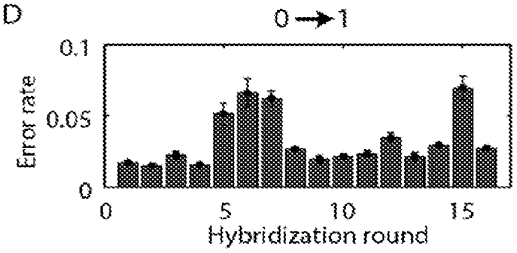


Fig. 16D

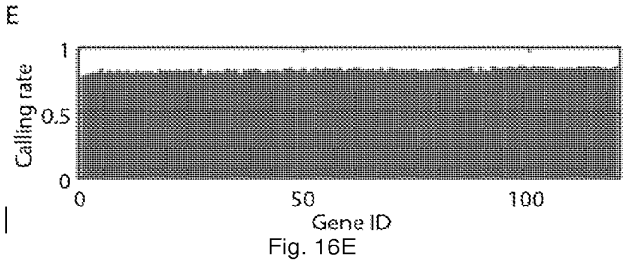


Fig. 16E

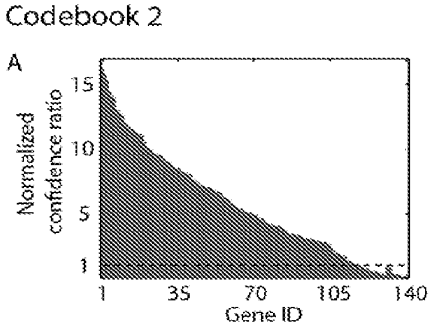


Fig. 17A

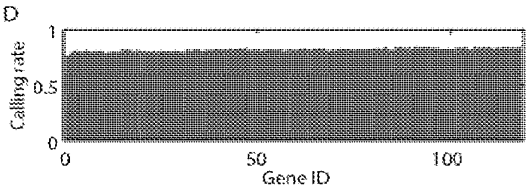


Fig. 17D

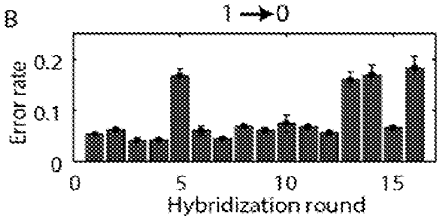


Fig. 17B

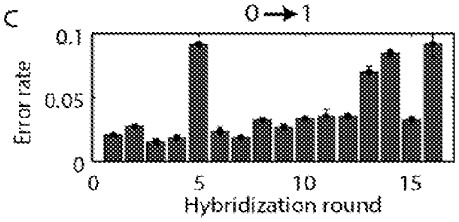


Fig. 17C

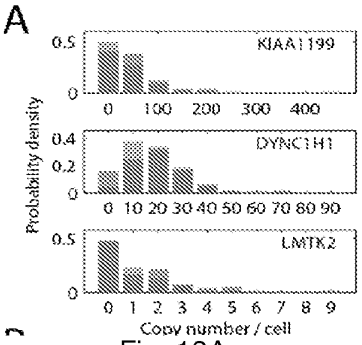


Fig. 18A

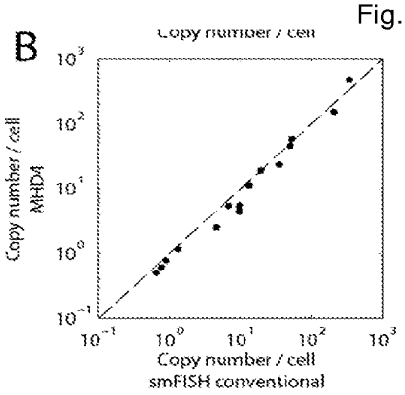


Fig. 18B

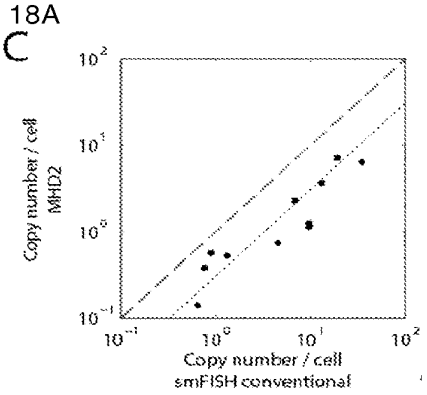


Fig. 18C

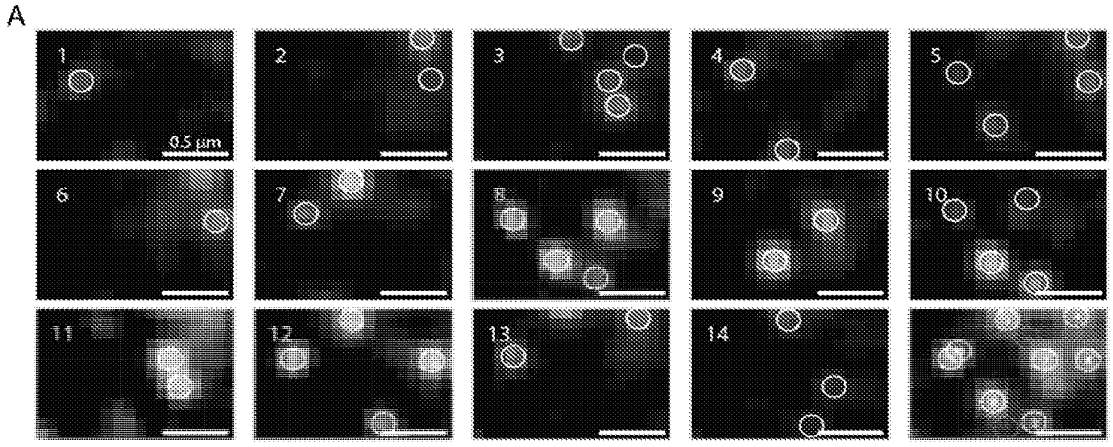


Fig. 19A

B

Spot number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Gene
1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	unidentified
2	0	0	0	0	0	0	0	1	0	1	0	1	0	1	FLNC
3	0	0	1	0	0	0	1	0	0	0	0	1	0	1	ZFPM2
4	0	1	0	0	1	1	0	0	0	0	0	0	1	0	LTBP2
5	1	0	0	0	0	0	0	1	0	0	0	1	1	0	TP53BP1
6	0	1	0	0	1	1	0	0	0	0	0	1	0	0	INTS1
7	0	0	1	0	0	0	0	1	1	0	1	0	0	0	PINK1-AS
8	0	0	0	0	1	0	0	1	1	1	0	0	0	0	MLLT4
9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	unidentified
10	0	0	0	1	1	0	1	0	0	1	0	0	0	0	TRIM14
11	0	0	0	1	0	0	0	0	0	0	0	0	0	0	unidentified
12	0	0	1	0	0	0	0	0	0	0	0	0	0	0	unidentified

Hybridization round

Fig. 19B

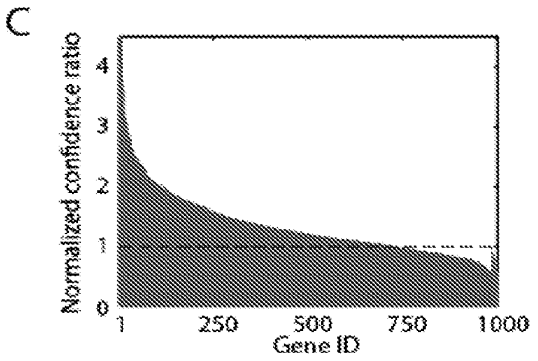


Fig. 19C

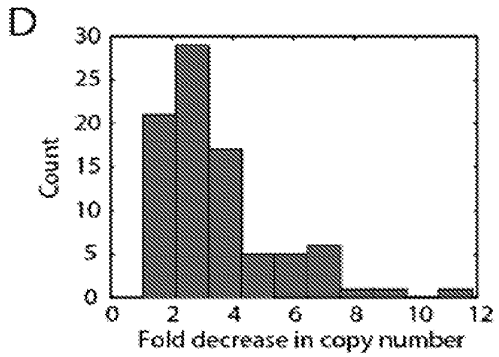


Fig. 19D

Fig. 20

Codebook 1 for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 1 0
FLNA	0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1
TNC	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
FBN2	0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0
KIAA1199	0 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0
COL5A1	0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
MALAT1	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
TLN1	0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0
FBN1	0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 0
SRRM2	0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 1
TEAD1	0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 0
IQGAP1	0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0
MYH10	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
PAPPA	0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0
NR4A1	0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
LRP1	0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0
SLC38A1	1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0
SMARCA5	0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1
CKAP5	0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0
PIEZO1	1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0
VCAN	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
PRRC2B	0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0
SON	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
SLC7A11	1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
GPR107	0 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0
KIAA1462	0 1 0 1 1 0 0 0 0 0 0 0 0 0 1 0
notarget001	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
blank001	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0 0
ROCK2	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
SPTBN1	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FLNC	0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0
PRPF8	0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1
IGF2R	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
FOSB	1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
UBR5	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
SVEP1	0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0
CBX5	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
PDS5A	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
NUMA1	0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1
PLXNA1	0 0 0 1 0 0 0 0 0 0 0 1 1 0 1 0
ALPK2	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0 0
AFAP1	0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0
SPTAN1	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
DOCK7	0 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0
NOTCH2	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
DYNC1H1	1 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0
ITGA2	1 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0
PRKDC	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
GTF3C1	0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0
THSD4	0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
PTPN14	0 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0
LUZP1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
AMOTL1	1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
MAN1A2	1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
notarget002	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
blank002	1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0
USP9X	0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0
CREBBP	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1

Codebook 2 (shuffled) for the 140-gene experiments

Genes	Codewords
THBS1	0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 1
FLNA	0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0
TNC	1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1
FBN2	0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1
KIAA1199	0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1
COL5A1	0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0
MALAT1	0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1
TLN1	0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 0
FBN1	1 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0
SRRM2	0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1
TEAD1	1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
IQGAP1	0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 1
MYH10	0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1
PAPPA	0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0
NR4A1	0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1
LRP1	0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0
SLC38A1	0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 1
SMARCA5	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1
CKAP5	0 0 1 0 0 1 0 0 0 0 0 0 0 1 1 0 0
PIEZO1	0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1
VCAN	0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0
PRRC2B	1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1
SON	0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0
SLC7A11	0 1 0 0 1 0 1 0 0 0 0 0 0 1 0 0
GPR107	0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 1
KIAA1462	0 0 1 1 1 0 0 0 0 0 0 0 1 0 0 0
notarget001	1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1
blank001	1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0
ROCK2	0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1
SPTBN1	0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 0
FLNC	0 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0
PRPF8	0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1
IGF2R	0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0
FOSB	0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 1
UBR5	0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0
SVEP1	0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0
CBX5	0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1
PDS5A	1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
NUMA1	0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0
PLXNA1	0 1 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0
ALPK2	0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0
AFAP1	1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0
SPTAN1	0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 0
DOCK7	0 0 1 0 0 0 0 0 0 0 0 1 1 0 1 0 0
NOTCH2	0 0 0 0 0 1 1 0 1 0 0 0 1 0 0 0
DYNC1H1	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0
ITGA2	0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0
PRKDC	0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 1
GTF3C1	0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 0
THSD4	1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
PTPN14	0 0 0 0 0 0 1 1 0 1 0 0 0 1 0 0
LUZP1	1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0
AMOTL1	0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0
MAN1A2	0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1
notarget002	1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0
blank002	1 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0
USP9X	0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0
CREBBP	0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 1

KPNA4	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
EGFR	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0
ZC3H13	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	
TNRC6A	0	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	
TPR	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0
COL7A1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	
ASCC3	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	
USP34	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1
CENPF	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	
KIDINS220	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	
SCUBE3	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	
NRIP1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	
NUP98	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	
ZFC3H1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	
FAM208B	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	
RAB3B	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	
AFF4	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	
PRKCA	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	
FLNB	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
CHD8	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
ANKRD52	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0
MKI67	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0
DIP2C	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	
SSH1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1
notarget003	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	
blank003	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	
DNAJC13	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	
DSEL	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	
PIK3CA	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	
MED14	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	
HERC2	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	
H6PD	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	
C14orf132	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0
SIPA1L3	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0
C17orf51	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	
NF1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	
VPS13A	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	
ABCA2	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	
PCNX	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	
NKTR	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	
FYCO1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	
ATP11A	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	
REV3L	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	
CASC5	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	
AKAP11	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	
CEP250	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1
USP24	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0
AGPS	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
CBL	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0
RNF169	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1
TMOD2	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0
FASN	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1
notarget004	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
blank004	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0
UBR2	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
PHIP	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1
GTF3C4	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0
QSER1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0
VPS13D	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0
FRY	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0
DIP2B	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0
HIVEP2	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0

KPNA4	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	
EGFR	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	
ZC3H13	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	
TNRC6A	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	
TPR	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	
COL7A1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	
ASCC3	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	
USP34	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	
CENPF	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	
KIDINS220	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1
SCUBE3	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1
NRIP1	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	
NUP98	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	
ZFC3H1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	
FAM208B	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	
RAB3B	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	
AFF4	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	
PRKCA	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	
FLNB	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
CHD8	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1
ANKRD52	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1
MKI67	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	
DIP2C	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	
SSH1	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	
notarget003	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
blank003	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0
DNAJC13	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	
DSEL	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	
PIK3CA	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0
MED14	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	
HERC2	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	
H6PD	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0
C14orf132	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	
SIPA1L3	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	
C17orf51	0	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	
NF1	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	
VPS13A	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	
ABCA2	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	
PCNX	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0	
NKTR	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	
FYCO1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	
ATP11A	1	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	
REV3L	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	
CASC5	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	
AKAP11	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	
CEP250	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	
USP24	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	
AGPS	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	
CBL	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	
RNF169	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	
TMOD2	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	
FASN	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	
notarget004	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	
blank004	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	
UBR2	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	
PHIP	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	
GTF3C4	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	
QSER1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	
VPS13D	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	
FRY	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	
DIP2B	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	
HIVEP2	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	

DOPEY1	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0
ALMS1	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0
HEATR5B	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0
HIVEP1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1
ARL10	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1
HELZ2	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1
POLQ	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
CACNA1A	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
KIAA2018	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1
LYST	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
ZKSCAN2	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1
DIEXF	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0
FZD4	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0
SLC5A3	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0
KIAA1147	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
LMTK2	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
STARD9	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0
BRCA2	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1
notarget005	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
blank005	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0

DOPEY1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0
ALMS1	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0
HEATR5B	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0
HIVEP1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0
ARL10	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
HELZ2	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0
POLQ	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0
CACNA1A	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0
KIAA2018	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0
LYST	0	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0
ZKSCAN2	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0
DIEXF	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0
FZD4	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0
SLC5A3	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
KIAA1147	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0
LMTK2	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0
STARD9	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0
BRCA2	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0
notarget005	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
blank005	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

DOCUMENT MADE AVAILABLE UNDER THE PATENT COOPERATION TREATY (PCT)

International application number:		**PCT/US2015/042556**
International filing date:		**29 July 2015 (29.07.2015)**
Document type:		**Certified copy of priority document**
Document details:	Country/Office:	**US**
	Number:	**62/142,653**
	Filing date:	**03 April 2015 (03.04.2015)**
Date of receipt at the International Bureau:		**03 August 2015 (03.08.2015)**

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a),(b) or (b-*bis*)

34, chemin des Colombettes
1211 Geneva 20, Switzerland
www.wipo.int

SEARCH HISTORY

Application Number	PCT/US 15/42556
Search Conducted By	DAM
Search Approved By	DAM

US/IPC/CPC Classifications Searched	IPC(8): C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01) CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 USPC: 506/9, 16; 536/24.3; 435/6.11
Date Conducted	16 November 2015

Documentation Searched	CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 + (text search) USPC: 506/9, 16; 536/24.33; 435/6.11 (text search)
Search Terms Used	nucleic acid probe, hybridization, barcode, index, next generation sequencing, read sequence, target sequence, error correction code, hamming code, fluorescence microscopy
Date Conducted	16 November 2015

Electronic Database Searched	PatBase
Files Searched	Full-text: AU BE BR CA CH CN DE DK EP ES FI FR GB IN JP KR SE TH TW US WO Bibliographic: (Europe) AT BA BE BG CH CS CY CZ DD DK EE ES FI GE GR HR HU IE IS IT LT LU LV MC MD MT NL NO PL PT RO RS SE SI SK SM TR UA YU (Asia) EA GC HK ID IL IN KZ MN MY PH RU SG SU TH TJ TW UZ VN (North America) CA CR CU DO GT HN MX NI PA SV TT (South America) AR BR CL CO EC PE UY (Australasia) AU NZ (Africa) AP DZ EG KE MA MW OA ZA ZM ZW
Date Conducted	16 November 2015

Search Logic:

Search 1: CPC=(B01J2219/00585 or B01J2219/00722 or C12Q1/6876) (Results 5390)
Search 2: IC=(C40B30/04 or C40B40/06 or C07H21/02 or C12Q1/68) (Results 100000)
Search 3: UC=(506/9 or 506/16 or 536/24.3 or 435/6.11) (Results 22061)
Search 4: 1 or 2 or 3 (Results 100000)
Search 5: 4 and FT=((barcode or indexing) and probe and (nucleic acid)) (Results 2129)
Search 6: 5 and FT=(mRNA) (Results 1351)
Search 7: 6 and FT=(fluorescen* microscopy) (Results 193)
Search 8: 5 and FT=((probe W4 (barcode or index or indicia)) and (probe W4 target%)) (Results 106)
Search 9: 8 and 7 (Results 18)
Search 10: 9 and FT=((Pool% or set%) W4 (probes)) (Results 5)
Search 11: 5 and FT=((error correction) and (hamming)) (Results 13)
Search 12: 11 and 7 (Results 2)
Search 13: 11 and 10 (Results 0)
Search 14: FT=((barcode or indexing) and probe and (nucleic acid)) (Results 3027)
Search 15: 14 and FT=(mRNA) (Results 1827)
Search 16: 15 and FT=(fluorescen* microscopy) (Results 234)
Search 17: 14 and FT=((probe W4 (barcode or index or indicia)) and (probe W4 target%)) (Results 122)
Search 18: 17 and 16 (Results 19)
Search 19: 14 and FT=((error correction) and (hamming)) (Results 13)
Search 20: 19 AND 16 (Results 2)
Search 21: 19 and 17 (Results 0)
Search 22: FT=(codeword) and INV=(ZHUANG Xiaowe) (Results 0)
Search 23: FT=(probe%) and INV=((Lubeck Eric) or (Cai Long)) (Results 6)

Electronic Database Searched	Google
Files Searched	Google Patents
Date Conducted	16 November 2015
Search Logic:	
probe hybridized to target and to barcode About 7,880 results (0.35 seconds) probe hybridized to target and to index sequence read About 72,600 results (0.54 seconds) multiplex fluorescent probes to adapter bar codes About 461 results (0.40 seconds) specific fluorescent probe for each barcode About 1,100 results (0.46 seconds) expose pool of barcoded probes to sample About 3,440 results (0.38 seconds) "pool of probes" added to samples Soderlund About 107 results (0.35 seconds) multiple probe sets exposed to cellular sample About 231,000 results (0.40 seconds) multiple nucleic acid probe sets exposed to cellular sample About 161,000 results (0.52 seconds)	

Electronic Database Searched	Google
Files Searched	Google Scholar
Date Conducted	16 November 2015
Search Logic:	
nucleic acid probe hybridized to target and to barcode About 11,600 results (**0.09** sec probe hybridized to target and to index sequence read About 47,100 results (**0.10** sec multiplex fluorescent probes to adapter bar codes About 4,940 results (**0.22** sec specific fluorescent probe for each barcode About 13,400 results (**0.06** sec expose pool of barcoded probes to sample About 4,110 results (**0.07** sec "pool of probes" barcoded added to sample About 47 results (**0.07** sec multiple nucleic acid probe sets exposed to cellular sample About 85,400 results (**0.17** sec	

```
============================================================
Sequence Listing was accepted.
See attached Validation Report.
If you need help call the Patent Electronic Business Center at (866)
217-9197 (toll free).
Reviewer: Douglas, Keisha (ASRC)
Timestamp: [year=2017; month=2; day=9; hr=16; min=45; sec=30; ms=86; ]
============================================================
```

Validated By CRFValidator v 1.0.5

Application No: 15329683 Version No: 1.0

Input Set:

Output Set:

Started: 2017-01-27 12:39:49.605
Finished: 2017-01-27 12:39:50.478
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 873 ms
Total Warnings: 40
Total Errors: 1
No. of SeqIDs Defined: 40
Actual SeqID Count: 40

Error code		Error Description
E	287	Invalid WIPO ST.2 date format; Use (YYYY-MM-DD) in <141>
W	213	Artificial or Unknown found in <213> in SEQ ID (1)
W	213	Artificial or Unknown found in <213> in SEQ ID (2)
W	213	Artificial or Unknown found in <213> in SEQ ID (3)
W	213	Artificial or Unknown found in <213> in SEQ ID (4)
W	213	Artificial or Unknown found in <213> in SEQ ID (5)
W	213	Artificial or Unknown found in <213> in SEQ ID (6)
W	213	Artificial or Unknown found in <213> in SEQ ID (7)
W	213	Artificial or Unknown found in <213> in SEQ ID (8)
W	213	Artificial or Unknown found in <213> in SEQ ID (9)
W	213	Artificial or Unknown found in <213> in SEQ ID (10)
W	213	Artificial or Unknown found in <213> in SEQ ID (11)
W	213	Artificial or Unknown found in <213> in SEQ ID (12)
W	213	Artificial or Unknown found in <213> in SEQ ID (13)
W	213	Artificial or Unknown found in <213> in SEQ ID (14)
W	213	Artificial or Unknown found in <213> in SEQ ID (15)
W	213	Artificial or Unknown found in <213> in SEQ ID (16)
W	213	Artificial or Unknown found in <213> in SEQ ID (17)
W	213	Artificial or Unknown found in <213> in SEQ ID (18)
W	213	Artificial or Unknown found in <213> in SEQ ID (19)

Input Set:

Output Set:

Started: 2017-01-27 12:39:49.605
Finished: 2017-01-27 12:39:50.478
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 873 ms
Total Warnings: 40
Total Errors: 1
No. of SeqIDs Defined: 40
Actual SeqID Count: 40

Error code	Error Description
W 213	Artificial or Unknown found in <213> in SEQ ID (20) This error has occured more than 20 times, will not be displayed

SEQUENCE LISTING

<110> President and Fellows of Harvard College

<120> SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

<130> H0498.70509US02

<140> US 15/329,683
<141> 2017-01-27

<150> PCT/US2015/042556
<151> 2015-07-29

<150> US 62/142,653
<151> 2015-04-03

<150> US 62/050,636
<151> 2014-09-15

<150> US 62/031,062
<151> 2014-07-30

<160> 40

<170> PatentIn version 3.5

<210> 1
<211> 120
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 1
gttggcgacg aaagcactgc gattggaacc gtcccaagcg ttgcgcttaa tggatcatca 60

attttgtctc actacgacgg tcaatcgcgc tgcatacttg cgtcggtcgg acaaacgagg 120

<210> 2
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 2
cgcaacgctt gggacggttc caatcggatc 30

<210> 3
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 3
cgaatgctct ggcctcgaac gaacgatagc 30

<210> 4
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 4
acaaatccga ccagatcgga cgatcatggg 30

<210> 5
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 5
caagtatgca gcgcgattga ccgtctcgtt 30

<210> 6
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 6
tgcgtcgtct ggctagcacg gcacgcaaat 30

<210> 7
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 7
aagtcgtacg ccgatgcgca gcaattcact 30

<210> 8

<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 8
cgaaacatcg gccacggtcc cgttgaactt 30

<210> 9
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 9
acgaatccac cgtccagcgc gtcaaacaga 30

<210> 10
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 10
cgcgaaatcc ccgtaacgag cgtcccttgc 30

<210> 11
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 11
gcatgagttg cctggcgttg cgacgactaa 30

<210> 12
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 12
ccgtcgtctc cggtccaccg ttgcgcttac 30

<210> 13
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 13
ggccaatggc ccaggtccgt cacgcaattt 30

<210> 14
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 14
ttgatcgaat cggagcgtag cggaatctgc 30

<210> 15
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 15
cgcgcggatc cgcttgtcgg gaacggatac 30

<210> 16
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 16
gcctcgatta cgacggatgt aattcggccg 30

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 17
gcccgtattc ccgcttgcga gtagggcaat 30

<210> 18
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 18
gttggtcggc acttgggtgc 20

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 19
cgatgcgcca attccggttc 20

<210> 20
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 20
cgcgggctat atgcgaaccg 20

<210> 21
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 21
taatacgact cactataggg aaagccggtt catccggtgg 40

<210> 22
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 22
taatacgact cactataggg tgatcatcgc tcgcgggttg 40

<210> 23
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 23
taatacgact cactataggg cgtggagggc atacaacgc 39

<210> 24
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 24
cgcaacgctt gggacggttc caatcggatc 30

<210> 25
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 25
cgaatgctct ggcctcgaac gaacgatagc 30

<210> 26

<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Guanine is modified by 3' Cy5 Sp

<400> 26
acaaatccga ccagatcgga cgatcatggg 30

<210> 27
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Thymine is modified by 3' Cy5 Sp

<400> 27
caagtatgca gcgcgattga ccgtctcgtt 30

<210> 28
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 28
gcgggaagca cgtggattag ggcatcgacc 30

<210> 29
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> (30)..(30)
<223> Thymine is modified by 3' Cy5 Sp

<220>
<223> Synthetic Polynucleotide

<400> 29
aagtcgtacg ccgatgcgca gcaattcact 30

<210> 30
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Thymine is modified by 3' Cy5 Sp

<400> 30
cgaaacatcg gccacggtcc cgttgaactt 30

<210> 31
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Adenine is modified by 3' Cy5 Sp

<400> 31
acgaatccac cgtccagcgc gtcaaacaga 30

<210> 32
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 32
cgcgaaatcc ccgtaacgag cgtcccttgc 30

<210> 33
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Adenine is modified by 3' Cy5 Sp

<400> 33
gcatgagttg cctggcgttg cgacgactaa 30

<210> 34
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 34
ccgtcgtctc cggtccaccg ttgcgcttac 30

<210> 35
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>

<221> misc_feature
<222> (30)..(30)
<223> Thymine is modified by 3' Cy5 Sp

<400> 35
ggccaatggc ccaggtccgt cacgcaattt 30

<210> 36
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 36
ttgatcgaat cggagcgtag cggaatctgc 30

<210> 37
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Cytosine is modified by 3' Cy5 Sp

<400> 37
cgcgcggatc cgcttgtcgg gaacggatac 30

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Guanine is modified by 3' Cy5 Sp

<400> 38
gcctcgatta cgacggatgt aattcggccg 30

<210> 39
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<220>
<221> misc_feature
<222> (30)..(30)
<223> Thymine is modified by 3' Cy5 Sp

<400> 39
gcccgtattc ccgcttgcga gtagggcaat 30

<210> 40
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Polynucleotide

<400> 40
cgggtttagc gccggaaatg 20

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
	FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet 1 of 2	GROUP ART UNIT: Not Yet Assigned	EXAMINER: Not Yet Assigned

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	U.S. Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		2004-0053300	A1	Soderlund et al.	03-18-2004
		2010-0291557	A1	Livak et al.	11-18-2010
		2010-0323348	A1	Hamady et al.	01-23-2010
		2010-0304994	A1	Wu et al.	12-02-2010
		2012-0035065	A1	Smolke et al.	02-09-2012
		2012-0129165	A1	Raj et al.	05-24-2012
		2013-0096014	A1	Anderson et al.	04-18-2013
		2013-0261019	A1	Lin et al.	10-03-2013
		2014-0031243	A1	Cai et al.	01-30-2014

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Foreign Patent Document Number	Foreign Patent Document Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Invitation to Pay Additional Fees mailed October 6, 2015 for Application No. PCT/US2015/042556 (H0498.70509WO00).	
		International Search Report and Written Opinion mailed December 18, 2015 for Application No. PCT/US2015/042556 (H0498.70509WO00).	
		International Preliminary Report on Patentability mailed February 9, 2017 for Application No. PCT/US2015/042556 (H0498.70509WO00).	
		International Search Report and Written Opinion mailed October 9, 2015 for Application No. PCT/US2015/042559 (H098.70510WO00).	
		International Preliminary Report on Patentability mailed February 9, 2017 for Application No. PCT/US2015/042559 (H098.70510WO00).	
		CHEN et al., Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015 Apr 24;348(6233):aaa6090. doi: 10.1126/science.aaa6090. Epub 2015 Apr 9.	
		FAKRUDDIN et al., Nucleic acid sequence based amplification (NASBA)-prospects and applications. Int. J. Life Sci. Pharm. Res. 2012 March; 2(1): L106-L121.	
		LUBECK et al., Single-cell in situ RNA profiling by sequential hybridization. Nature Methods. 2014 Apr; 11(4): 360-1.	
		LUBECK et al., Single-cell in situ RNA profiling by sequential hybridization. Nature Methods. 2014 Apr; 11(4): 360-1. Supplementary Information.	

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	2	of	2	GROUP ART UNIT: Not Yet Assigned	EXAMINER: Not Yet Assigned

		MOFFITT et al., High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc Natl Acad Sci U S A. 2016 Dec 13;113(50):14456-14461. Epub 2016 Nov 22.	
		MOFFITT et al., High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci U S A. 2016 Sep 27;113(39):11046-51. doi: 10.1073/pnas.1612826113. Epub 2016 Sep 13.	
		MOFFITT et al., RNA Imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH). Methods Enzymol. 2016;572:1-49. doi: 10.1016/bs.mie.2016.03.020. Epub 2016 April 27.	
		SHIELS et al., RNA-DNA hybrids containing damaged DNA are substrates for RNase H. Bioorg Med Chem Lett. 2001 Oct 8;11(19):2623-6.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

(PCT Article 17(3)(a) and Rules 40.1 and 40.2(e))

Applicant's or agent's file reference H0498.70509	Date of mailing *(day/month/year)* 06 OCT 2015 **PAYMENT DUE** within **ONE MONTH** from the above date of mailing
International application No. PCT/US 15/42556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. This International Searching Authority

(i) considers that there are 4 *(number of)* inventions claimed in the international application covered by the claims indicated below/on an extra sheet:
----Go to Extra Sheet for continuation----

(ii) therefore considers that **the international application does not comply with the requirement of unity of invention** (Rules 13.1, 13.2 and 13.3) for the reasons indicated below/on an extra sheet:
----Go to Extra Sheet for continuation----

(iii) ☐ has carried out a partial international search (see Annex) ☒ will establish the international search report on those parts of the international application which relate to the invention first mentioned in claims Nos.:
Claims 1-3, 88-90, 166-171

(iv) will establish the international search report on the other parts of the international application only if, and to the extent to which, additional fees are paid.

2. Consequently, the applicant is hereby **invited to pay**, within the time limit indicated above, additional fees in the amount indicated below:

$2,080.00	x	3	=	$6,240.00 (See Item 2 of annex)
Fee per additional invention		number of additional inventions		total amount of additional fees/currency

3. The applicant is informed that, according to Rule 40.2(c), **the payment of any additional fees may be made under protest**, that is, a reasoned statement to the effect that the international application complies with the requirement of unity of invention or that the amount of the required additional fees is excessive, where applicable, subject to the payment of a protest fee.

Where the applicant pays additional fees under protest, the applicant is hereby invited, within the time limit indicated above, to pay a protest fee (Rule 40.2(e)) in the amount of _____ *(amount/currency)*

Where the applicant has not, within the time limit indicated above, paid the required protest fee, the protest will be considered not to have been made and the International Searching Authority will so declare.

4. ☒ Claim(s) Nos. 4-87,99-121,125-135,142,143,148-165,172-176 have been found to be unsearchable under Article 17(2)(b) because of defects under Article 17(2)(a) and therefore have not been included with any invention.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300

Form PCT/ISA/206 (April 2005)

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US 15/42556

Item 2 (continued). For International Applications filed on or after 01 January 2014, Applicant is reminded that the search fee per additional invention indicated in item 2 is the undiscounted fee per additional invention. An Applicant may pay the search fee per additional invention fee reduced by 50% (small entity assertion) or 75% (micro entity certification), as appropriate. See 37 CFR 1.27 and 1.29.

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.

4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 AI to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 AI to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore- labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").
-----continued on next sheet-----

Form PCT/ISA/206 (extra sheet) (April 2005)

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US 15/42556

---continued from previous sheet--------

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

Form PCT/ISA/206 (extra sheet) (April 2005)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Date of mailing *(day/month/year)*	18 DEC 2015
Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 15/42556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **19 months from the priority date, the applicant may request that a supplementary internatioual search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee W. Young PCT Helpdesk: 571-272-4300 Telephone No. PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2014)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 15/42556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)	(Earliest) Priority Date *(day/month/year)* 30 July 2014 (30.07.2014)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 9 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

   c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☒ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

   a. the figure of the **drawings** to be published with the abstract is Figure No. 1B

   ☒ as suggested by the applicant.

   ☐ as selected by this Authority, because the applicant failed to suggest a figure.

   ☐ as selected by this Authority, because this figure better characterizes the invention.

   b. ☐ none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

Box No. I Nncleotide and/or amino acid seqnence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-87,99-121,125-135,142,143,148-165,172-176
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
----Go to Extra Sheet for Continuation-----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Claims 1-3, 88-90, 166-171

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 (text search)
USPC: 506/9, 16; 536/24.33; 435/6.11 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Patents; Google Scholar
Search terms: nucleic acid probe, hybridization, barcode, index, next generation sequencing, read sequence, target sequence, error correction code, hamming code, fluorescence microscopy, transcriptsome

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/0031243 A1 (CAI et al.) 30 January 2014 (30.01.2014). Especially para [0007], [0018], [0052], [0069], [0141], [0178]	1-3, 88-90,166-171
Y	US 2010/0323348 A1 (HAMADY et al.) 23 January 2010 (23.01.2010).Especially para [0016], [0098].	1-3, 90, 166-171
Y	US 2010/0291557 A1 (LIVAK et al.) 18 November 2010 (18.11.2010). Especially para [0062], [0063]	88-90, 166-171
Y	US 2004/0053300 A1 (SODERLUND et al.) 18 March 2004 (18.03.2004). Especially para [0033], [0035], [0049], [0053]	166-171
X,P	CHEN et al. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science ePub 9 April 2015 Vol 348 No 6233 Pages 1-14. entire article.	1-3, 88-90,166-171

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 November 2015 (16.11.2015)	18 DEC 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42556

----Continuation of Box III (Lack of Unity of Invention)------

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.

4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").
-----continued on next sheet-----

----continued from previous sheet---------

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe -specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	18 DEC 2015
Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 15/42556	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule *43bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 *bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	16 November 2015 (16.11.2015)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. --see below---- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

   ☐ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.
4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.
------continued in Supplemental Box-----

4. Consequently, this opinion has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. Claims 1-3, 88-90, 166-171

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	1-3, 88-90, 166-171	NO
Industrial applicability (IA)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO

2. Citations and explanations:

Claims 1-3 lack an inventive step under PCT Article 33(3) as being obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai"), in view of US 2010/0323348 A1 to HAMADY et al. (hereinafter "Hamady").

As to claim 1, Cai teaches exposing a sample to a plurality of nucleic acid probes (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode"), for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample (para [0007]); creating codewords based on the binding of the nucleic acid probes (para [0007]) and for at least some of the codewords, matching the codeword to a valid codeword wherein (para [0007]). Cai does not teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches applying error correction to the codeword to form a valid codeword (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; "Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected"). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine all claim limitations, with the exception of if no match is found, applying error correction to the codeword to form a valid codeword, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled characterizing multiple isoforms of mRNA.

As to claim 2, Cai teaches exposing the sample to at least 5 different nucleic acid probes (para [0052]; "comprises five oligo nucleotide probes").

As to claim 3, Cai further teaches exposing the sample to at least 10 different nucleic acid probes (para [0069]; 8 probes associated with a label emitting a green (G) signal; and 2 probes associated with a label emitting a blue (B) signal").

-----continued in Supplemental Box-----

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**

Continuation of:
Box IV-3 (Lack of Unity of Invention)

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore- labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 (Citations and Explanations)

Claims 88, 89 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of US 2010/0291557 A1 to LIVAK et al. (hereinafter "Livak").

As to claim 88, Cai teaches method, comprising: exposing a sample to a plurality of nucleic acid probes (para [0007]), and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes (para [0007]) formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences (para [0178]; "Super-resolution barcoding provides a useful single cell follow-up technique to high-throughput sequencing technologies"). Cai does not teach the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample. However, Livak teaches the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences (para [0062-0063]; "the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes specific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay") and for each of the nucleic acid probes determining binding of the target sequences of the nucleic acid and probes within the sample (para [0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Consequently, it would have been obvious for one of ordinary skill in the art to combine exposing a sample to a plurality of nucleic acid probes, at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences, all as taught by Cai, with the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, as taught by Livak, with determining binding of the target sequences of the nucleic acid and probes within the sample, also as taught by Livak, because it would have enabled a method of identifying and quantitating a mRNA population in a cell or group of cells.

As to claim 89, Cai further teaches creating codewords based on determination of the read sequences within the sample (para [0007], [0178]).

Claim 90 lacks an inventive step under PCT Article 33(3) as being obvious over Cai and Livak, in view of Hamady.

As to claim 90, Cai teaches codewords (para [0007]) and matching the codeword to a valid codeword (para [0007]). Neither Cai nor Livak teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches if no match is found, applying error correction to the codeword to form a valid codeword (para [0016], [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine codewords and matching the codeword to a valid codeword, as taught by Cai, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled because it would have enabled characterizing multiple isoforms of mRNA.

--------continued in next Supplemental Box-------

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 and previous Supplemental Box

Claims 166-171 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of Hamady, in further view of Livak, in final view of US 2004/0053300 A1 to SODERLUND et al (hereinafter "Soderlund").

As to claim 166, Cai teaches a method, comprising: associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords (para [0007]) wherein the codewords comprise a number of positions and values for each position (para [0007]). Cai does not teach the codewords form an error-checking and/or error-correcting code space. However, Hamady teaches that the codewords form an error-checking and/or error-correcting code space (para [0016]; [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Cai further teaches forming a plurality of nucleic acid probes (para [0007]) but does not teach that the plurality of probes comprise a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets. However, Livak teaches probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets (para [0062-0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Although Cai further teaches each of the probe pools corresponds to a specific value of a unique position within the codewords (para [0007]; [0162]- In experiments related to those previously described herein, the inventors targeted the YLR414c mRNA with3 sets of oligo probes labeled with different fluorophores"), Cai does not teach grouping the nucleic acid probes into a plurality of probe pools, exposing a sample to one of the probe pools; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions. However, Soderlund teaches grouping the nucleic acid probes into a plurality of probe pools (para [0049]; One or more soluble pools are provided with preset, but optional numbers, preferably varying between 10-500, more preferably between 50-400, most preferably between 100-300 soluble polynucleotide sequences, with distinct sizes allowing their accurate identification, e.g. by mass spectrometry. The soluble probes, which can be identified without any tracer tags, can alternatively be provided with tags"; para [0044]-" Preferred tracer tags are fluorochromes or fluorophors, especially such which have different wave-lengths of emission";. para [0033]-" The term "pool" means a subset or a library of soluble or solubilizable polynucleotide probes. Each pool comprises an optional defined number of polynucleotide probes. A convenient optional number is, for example, approximately 10 probes"; exposing a sample to one of the probe pools (para [0053]; In the basic embodiment of the present invention tracer or primer tagged DNA probes are allowed to hybridize with the preparation of RNA, which is to be analyzed. The analyte polynucleotide sequences or RNA analytes, present in the cell or tissue sample to be determined, are isolated by per se known methods"); Soderlund further teaches repeating the exposing step one or more times, using a different probe pool for at least some of the repetitions (para [0035]; However, because the nucleotide fragments are not characterized, some of the nucleotide fragments obtained by cleaving with various restriction enzymes may by chance have identical sizes. This means that the probes are only partly identifiable by their sizes. When repeating the tests, the redundancy is soon detected and such probes, which cannot be identified with 100% accuracy may be deleted or replaced with other more convenient probes"). Cai further teaches imaging a sample of probes (para [0141]; Imaging of the hybridized cells is carried out on automated fluorescence microscopes"). An artisan of ordinary skill in the art would have recognized that using pools or sets of probes would have enabled obtaining large amounts of information about a given sample, particularly if different probe pools were used in repetitive analysis, as taught by Soderlund. Consequently, it would have been obvious for one of ordinary skill in the art to combine associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, as taught by Cai, with the codewords form an error-checking and/or error-correcting code space, as taught by Hamady, with probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets, as taught by Livak, with grouping the nucleic acid probes into a plurality of probe pools, as taught by Soderlund, with exposing a sample to one of the probe pools; and repeating the exposing one or more times, using a different probe pool for at least some of the repetitions, again as taught by Soderlund, with using imaging technology to visualize the probes, as taught by Cai, because it would have enabled providing an expression profile of a sample.

As to claim 167, Cai further teaches combining the images of the sample into a combined image (para [0165]; "After many iterations of this cycle, a high resolution image can be reconstructed from all of the centroid fits in the individual images").

As to claim 168, Cai further teaches identifying foreground features corresponding to putative nucleic acid targets in the sample (para [0018]; enabling robust hybridization and identification by unique combinations of fluorophores").

As to claim 169, Cai further teaches determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample (para [0007], [0018], [0165]).

As to claim 170, Cai further teaches matching the determined codeword to a valid codeword wherein (para [0007]). Additionally, Hamady further teaches if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target (para [0016], [0098]).

As to claim 171, Cai further teaches imaging the sample using fluorescence microscopy (para [0141]; "Imaging of the hybridized cells is carried out on automated fluorescence microscopes").

Claims 1-3, 88-90, 166-171 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

SEARCH HISTORY

Application Number	PCT/US 15/42556
Search Conducted By	DAM
Search Approved By	DAM

US/IPC/CPC Classifications Searched	IPC(8): C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01) CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 USPC: 506/9, 16; 536/24.3; 435/6.11
Date Conducted	16 November 2015

Documentation Searched	CPC: B01J 2219/00585, 2219/00722; C12Q 1/6876 + (text search) USPC: 506/9, 16; 536/24.33; 435/6.11 (text search)
Search Terms Used	nucleic acid probe, hybridization, barcode, index, next generation sequencing, read sequence, target sequence, error correction code, hamming code, fluorescence microscopy
Date Conducted	16 November 2015

Electronic Database Searched	PatBase
Files Searched	Full-text: AU BE BR CA CH CN DE DK EP ES FI FR GB IN JP KR SE TH TW US WO Bibliographic: (Europe) AT BA BE BG CH CS CY CZ DD DK EE ES FI GE GR HR HU IE IS IT LT LU LV MC MD MT NL NO PL PT RO RS SE SI SK SM TR UA YU (Asia) EA GC HK ID IL IN KZ MN MY PH RU SG SU TH TJ TW UZ VN (North America) CA CR CU DO GT HN MX NI PA SV TT (South America) AR BR CL CO EC PE UY (Australasia) AU NZ (Africa) AP DZ EG KE MA MW OA ZA ZM ZW
Date Conducted	16 November 2015

Search Logic:

Search 1: CPC=(B01J2219/00585 or B01J2219/00722 or C12Q1/6876) (Results 5390)
Search 2: IC=(C40B30/04 or C40B40/06 or C07H21/02 or C12Q1/68) (Results 100000)
Search 3: UC=(506/9 or 506/16 or 536/24.3 or 435/6.11) (Results 22061)
Search 4: 1 or 2 or 3 (Results 100000)
Search 5: 4 and FT=((barcode or indexing) and probe and (nucleic acid)) (Results 2129)
Search 6: 5 and FT=(mRNA) (Results 1351)
Search 7: 6 and FT=(fluorescen* microscopy) (Results 193)
Search 8: 5 and FT=((probe W4 (barcode or index or indicia)) and (probe W4 target%)) (Results 106)
Search 9: 8 and 7 (Results 18)
Search 10: 9 and FT=((Pool% or set%) W4 (probes)) (Results 5)
Search 11: 5 and FT=((error correction) and (hamming)) (Results 13)
Search 12: 11 and 7 (Results 2)
Search 13: 11 and 10 (Results 0)
Search 14: FT=((barcode or indexing) and probe and (nucleic acid)) (Results 3027)
Search 15: 14 and FT=(mRNA) (Results 1827)
Search 16: 15 and FT=(fluorescen* microscopy) (Results 234)
Search 17: 14 and FT=((probe W4 (barcode or index or indicia)) and (probe W4 target%)) (Results 122)
Search 18: 17 and 16 (Results 19)
Search 19: 14 and FT=((error correction) and (hamming)) (Results 13)
Search 20: 19 AND 16 (Results 2)
Search 21: 19 and 17 (Results 0)
Search 22: FT=(codeword) and INV=(ZHUANG Xiaowe) (Results 0)
Search 23: FT=(probe%) and INV=((Lubeck Eric) or (Cai Long)) (Results 6)

Electronic Database Searched	Google
Files Searched	Google Patents
Date Conducted	16 November 2015

Search Logic:

probe hybridized to target and to barcode About 7,880 results (0.35 seconds)
probe hybridized to target and to index sequence read About 72,600 results (0.54 seconds)
multiplex fluorescent probes to adapter bar codes About 461 results (0.40 seconds)
specific fluorescent probe for each barcode About 1,100 results (0.46 seconds)
expose pool of barcoded probes to sample About 3,440 results (0.38 seconds)
"pool of probes" added to samples Soderlund About 107 results (0.35 seconds)
multiple probe sets exposed to cellular sample About 231,000 results (0.40 seconds)
multiple nucleic acid probe sets exposed to cellular sample About 161,000 results (0.52 seconds)

Electronic Database Searched	Google
Files Searched	Google Scholar
Date Conducted	16 November 2015

Search Logic:

nucleic acid probe hybridized to target and to barcode About 11,600 results (**0.09** sec
probe hybridized to target and to index sequence read About 47,100 results (**0.10** sec
multiplex fluorescent probes to adapter bar codes About 4,940 results (**0.22** sec
specific fluorescent probe for each barcode About 13,400 results (**0.06** sec
expose pool of barcoded probes to sample About 4,110 results (**0.07** sec
"pool of probes" barcoded added to sample About 47 results (**0.07** sec
multiple nucleic acid probe sets exposed to cellular sample About 85,400 results (**0.17** sec

PATENT COOPERATION TREATY

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 09 February 2017 (09.02.2017)	
Applicant's or agent's file reference H0498.70509	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2015/042556	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Mineko Mohri

e-mail: pct.team8@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2015/042556	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)	Priority date *(day/month/year)* 30 July 2014 (30.07.2014)
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 10 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☒	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☒	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☒	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 31 January 2017 (31.01.2017)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Mineko Mohri e-mail: pct.team8@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	18 DEC 2015
Applicant's or agent's file reference H0498.70509	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 15/42556	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C40B 30/04, 40/06; C07H 21/02; C12Q 1/68 (2015.01)
CPC - B01J 2219/00585, 2219/00722; C12Q 1/6876

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule *43bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 *bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	16 November 2015 (16.11.2015)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. --see below---- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-87,99-121,125-135,142,143,148-165,172-176

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

   ☐ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-3, 88-90, 166-171, drawn to a method of exposing a plurality of (primary) nucleic acid probes to a sample.

Group II: Claims 91-94, 144-147, drawn to a method of exposing a sample containing a plurality of (primary) nucleic acid probes to a plurality of secondary nucleic acid probes that are complementary to a segment of the primary nucleic acid probes and have a signaling entity attached.

Group III: Claims 95-98, drawn to a composition of a plurality of primary nucleic acid probes.

Group IV: Claims 122-124, 136-141, drawn to a method of forming a nucleic acid probe comprising a target sequence and one or more read sequences involving associating a plurality of targets with a plurality of target sequences and a plurality of codewords and associating a plurality of distinguishable read sequences with the plurality of codewords.

The inventions listed as Groups I-IV do not t relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Groups I and II have the special technical feature of exposing a sample to nucleic acid probes, not required by Groups III or IV.

Group I has the special technical feature of creating codewords based on the binding of the primary nucleic acid probes, not required by Group II-IV.

Group II has the special technical feature of creating codewords based on the binding of the secondary nucleic acid probes, not required by Group I, III or IV.

Group III has the special technical feature of being a composition of a plurality of primary nucleic acid probes, not required by Groups I, II, or IV.

Group IV has the special technical feature of a method of designing a primary nucleic acid probe, not required by Groups I, II or III.

Common Technical Features:

1. Groups I-IV share the common technical feature of a plurality of primary nucleic acid probes that comprise codewords.

2. Groups I-IV share the common technical feature of error-correction of a codeword(s).

3. Groups I-IV share the common technical feature of a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences.
4. Groups I and II share the common technical feature of exposing sample to a plurality of nucleic acid probe.
-----continued in Supplemental Box-----

4. Consequently, this opinion has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. Claims 1-3, 88-90, 166-171

Form PCT/ISA/237 (Box No. IV) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	1-3, 88-90, 166-171	NO
Industrial applicability (IA)	Claims	1-3, 88-90, 166-171	YES
	Claims	NONE	NO

2. Citations and explanations:

Claims 1-3 lack an inventive step under PCT Article 33(3) as being obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai"), in view of US 2010/0323348 A1 to HAMADY et al. (hereinafter "Hamady").

As to claim 1, Cai teaches exposing a sample to a plurality of nucleic acid probes (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore-labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode"), for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample (para [0007]); creating codewords based on the binding of the nucleic acid probes (para [0007]) and for at least some of the codewords, matching the codeword to a valid codeword wherein (para [0007]). Cai does not teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches applying error correction to the codeword to form a valid codeword (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; "Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected"). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine all claim limitations, with the exception of if no match is found, applying error correction to the codeword to form a valid codeword, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled characterizing multiple isoforms of mRNA.

As to claim 2, Cai teaches exposing the sample to at least 5 different nucleic acid probes (para [0052]; "comprises five oligo nucleotide probes").

As to claim 3, Cai further teaches exposing the sample to at least 10 different nucleic acid probes (para [0069]; 8 probes associated with a label emitting a green (G) signal; and 2 probes associated with a label emitting a blue (B) signal").

-----continued in Supplemental Box-----

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
Box IV-3 (Lack of Unity of Invention)

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2014/0031243 A1 to CAI et al. (hereinafter "Cai") [published 30 January 2014], in view of US 2010/0323348 Al to HAMADY et al. (hereinafter "Hamady"), in further view of in view of US 2013/0261019 A1 to LIN et al. (hereinafter "Lin") and in further view of US 2010/0291557 Al to LIVAK et al. (hereinafter "Livak").

Concerning common technical feature #1 and #4, a plurality of primary nucleic acid probes that comprise codewords [e.g. 2 or more isoforms of a mRNA], and exposing a sample to a plurality of nucleic acid probes, Cai teaches (para [0007]; "In various embodiments, the invention teaches a method for characterizing one or more mRNA isoforms in a single cell, including: (i) creating a molecular barcode for each of said one or more mRNA isoforms in the cell, comprising: (a) providing two or more sets of two or more fluorophore- labeled oligonucleotide probes, wherein each set of probes are configured to hybridize with a specific region of one or more mRNA isoforms; and (b) hybridizing, within said cell, a quantity of said one or more mRNA isoforms with a quantity of said probes specific thereto, wherein each mRNA isoform that is hybridized with said fluorophore-labeled oligonucleotide probes emits two or more distinct signals, so as to create the molecular barcode; and (ii) resolving the molecular barcode").

Concerning common technical feature #2, error-correction of a codeword(s), error-correction of a codeword(s) based on, for example, different isoforms of a mRNA generating different or incomplete signals, would have been obvious based on the teaching of Cai, as set forth in common technical feature #1. Furthermore, Hamady teaches (para [0016]; The term "Hamming code" as used herein, refers an arithmetic process that identifies unique binary codes based upon inherent redundancy that are capable of correcting single bit errors. For example, a Hamming code can be matched with a nucleic acid barcode in order to screen for single nucleotide errors occurring during nucleic acid amplification. The identification of a single nucleotide error by using a Hamming code, thereby allows for the correction of the nucleic acid barcode"; para [0098]; Hamming codes may be represented by a subset of the possible codewords that are chosen from the center of multidimensional spheres (i.e., for example, hyperspheres) in a binary subspace. Single bit errors may fall within hyperspheres associated with a specific codeword and can thus be corrected").

Concerning common technical feature #3, a primary nucleic acid probe comprising a target binding sequence and one or more artificial read sequences, Lin teaches (para [0004]; The invention provides, inter alia, nucleic acid barcode probes that comprise a nucleic acid [....] The barcode probes may further comprise a target binding moiety"). Furthermore, concerning secondary probes to read a probe, Livak teaches (para [0062]) the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes pecific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature

Note concerning item 4: Claims 4-87, 99-121, 125-135, 142, 143, 148-165, 172-176 are multiple dependent claims and are not drafted according to the second and third sentences of PCT Rule 6.4(a).

Note: claim 51 cannot depend on itself. For the purpose of this invitation, claim 51 is interpreted as a dependent claim of claim 50.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 (Citations and Explanations)

Claims 88, 89 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of US 2010/0291557 A1 to LIVAK et al. (hereinafter "Livak").

As to claim 88, Cai teaches method, comprising: exposing a sample to a plurality of nucleic acid probes (para [0007]), and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes (para [0007]) formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences (para [0178]; "Super-resolution barcoding provides a useful single cell follow-up technique to high-throughput sequencing technologies"). Cai does not teach the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample. However, Livak teaches the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences (para [0062-0063]; "the invention comprises a binary probe and clamp composition in which the probe comprises a target-specific portion and a clamp-specific portion. The target- specific portion is capable of sequence-specific binding to a target polynucleotide sequence. A clamp comprises a probes specific portion and a label. The clamp-specific portion of the probe and the probe-specific portion of the clamp form a duplex structure in the binary probe and clamp composition .... [0063] During a nucleic acid hybridization assay, the target polynucleotide sequence may be detected and quantitated in the sample by detection of the label on the binary composition. The clamp-specific portion of the probe is bound to the probe-specific portion of the clamp during the detection phase of the assay") and for each of the nucleic acid probes determining binding of the target sequences of the nucleic acid and probes within the sample (para [0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Consequently, it would have been obvious for one of ordinary skill in the art to combine exposing a sample to a plurality of nucleic acid probes, at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences, all as taught by Cai, with the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences and for each of the nucleic acid probes, as taught by Livak, with determining binding of the target sequences of the nucleic acid and probes within the sample, also as taught by Livak, because it would have enabled a method of identifying and quantitating a mRNA population in a cell or group of cells.

As to claim 89, Cai further teaches creating codewords based on determination of the read sequences within the sample (para [0007], [0178]).

Claim 90 lacks an inventive step under PCT Article 33(3) as being obvious over Cai and Livak, in view of Hamady.

As to claim 90, Cai teaches codewords (para [0007]) and matching the codeword to a valid codeword (para [0007]). Neither Cai nor Livak teach if no match is found, applying error correction to the codeword to form a valid codeword. However, Hamady teaches if no match is found, applying error correction to the codeword to form a valid codeword (para [0016], [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Consequently, it would have been obvious for one of ordinary skill in the art to combine codewords and matching the codeword to a valid codeword, as taught by Cai, with if no match is found, applying error correction to the codeword to form a valid codeword, as taught by Hamady, because it would have enabled because it would have enabled characterizing multiple isoforms of mRNA.

--------continued in next Supplemental Box-------

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42556

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box V-2 and previous Supplemental Box

Claims 166-171 lack an inventive step under PCT Article 33(3) as being obvious over Cai, in view of Hamady, in further view of Livak, in final view of US 2004/0053300 A1 to SODERLUND et al (hereinafter "Soderlund").

As to claim 166, Cai teaches a method, comprising: associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords (para [0007]) wherein the codewords comprise a number of positions and values for each position (para [0007]). Cai does not teach the codewords form an error-checking and/or error-correcting code space. However, Hamady teaches that the codewords form an error-checking and/or error-correcting code space (para [0016]; [0098]). An artisan of ordinary skill in the art would have recognized that using Hamming codes, as taught by Hamady, would have enabled correction of a misread of a codeword or creating a new codeword for an unknown or uncharacterized mRNA isoform, and thus would have been useful in the method taught by Cai. Cai further teaches forming a plurality of nucleic acid probes (para [0007]) but does not teach that the plurality of probes comprise a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets. However, Livak teaches probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets (para [0062-0063]). An artisan of ordinary skill in the art would have recognized that using a probe comprising a target specific portion and a read specific barcode would have enabled a method to assign a specific target to a specific cell or group of cells for instance, during a multiplex sequence read. Although Cai further teaches each of the probe pools corresponds to a specific value of a unique position within the codewords (para [0007]; [0162]- In experiments related to those previously described herein, the inventors targeted the YLR414c mRNA with3 sets of oligo probes labeled with different fluorophores"), Cai does not teach grouping the nucleic acid probes into a plurality of probe pools, exposing a sample to one of the probe pools; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions. However, Soderlund teaches grouping the nucleic acid probes into a plurality of probe pools (para [0049]; One or more soluble pools are provided with preset, but optional numbers, preferably varying between 10-500, more preferably between 50-400, most preferably between 100-300 soluble polynucleotide sequences, with distinct sizes allowing their accurate identification, e.g. by mass spectrometry. The soluble probes, which can be identified without any tracer tags, can alternatively be provided with tags"; para [0044]-" Preferred tracer tags are fluorochromes or fluorophors, especially such which have different wave-lengths of emission";. para [0033]-" The term "pool" means a subset or a library of soluble or solubilizable polynucleotide probes. Each pool comprises an optional defined number of polynucleotide probes. A convenient optional number is, for example, approximately 10 probes"; exposing a sample to one of the probe pools (para [0053]; In the basic embodiment of the present invention tracer or primer tagged DNA probes are allowed to hybridize with the preparation of RNA, which is to be analyzed. The analyte polynucleotide sequences or RNA analytes, present in the cell or tissue sample to be determined, are isolated by per se known methods"); Soderlund further teaches repeating the exposing step one or more times, using a different probe pool for at least some of the repetitions (para [0035]; However, because the nucleotide fragments are not characterized, some of the nucleotide fragments obtained by cleaving with various restriction enzymes may by chance have identical sizes. This means that the probes are only partly identifiable by their sizes. When repeating the tests, the redundancy is soon detected and such probes, which cannot be identified with 100% accuracy may be deleted or replaced with other more convenient probes"). Cai further teaches imaging a sample of probes (para [0141]; Imaging of the hybridized cells is carried out on automated fluorescence microscopes"). An artisan of ordinary skill in the art would have recognized that using pools or sets of probes would have enabled obtaining large amounts of information about a given sample, particularly if different probe pools were used in repetitive analysis, as taught by Soderlund. Consequently, it would have been obvious for one of ordinary skill in the art to combine associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, as taught by Cai, with the codewords form an error-checking and/or error-correcting code space, as taught by Hamady, with probes that comprise signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets, as taught by Livak, with grouping the nucleic acid probes into a plurality of probe pools, as taught by Soderlund, with exposing a sample to one of the probe pools; and repeating the exposing one or more times, using a different probe pool for at least some of the repetitions, again as taught by Soderlund, with using imaging technology to visualize the probes, as taught by Cai, because it would have enabled providing an expression profile of a sample.

As to claim 167, Cai further teaches combining the images of the sample into a combined image (para [0165]; "After many iterations of this cycle, a high resolution image can be reconstructed from all of the centroid fits in the individual images").

As to claim 168, Cai further teaches identifying foreground features corresponding to putative nucleic acid targets in the sample (para [0018]; enabling robust hybridization and identification by unique combinations of fluorophores").

As to claim 169, Cai further teaches determining the codeword corresponding to each foreground feature by determining binding patterns of the nucleic acid probe pools to the sample (para [0007], [0018], [0165]).

As to claim 170, Cai further teaches matching the determined codeword to a valid codeword wherein (para [0007]). Additionally, Hamady further teaches if no match is found, applying error correction to the determined codeword to form a valid codeword that correspond to a nucleic acid target (para [0016], [0098]).

As to claim 171, Cai further teaches imaging the sample using fluorescence microscopy (para [0141]; "Imaging of the hybridized cells is carried out on automated fluorescence microscopes").

Claims 1-3, 88-90, 166-171 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 09 OCT 2015
Applicant's or agent's file reference H0498.70510	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 15/42559	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **19 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee W. Young Telephone No. PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2014)

RECEIVED
Wolf, Greenfield & Sacks, P.C.
OCT 16 2015
Docketed ✓ Already Docketed
Not Required
Initials 1st 2nd

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference H0498.70510	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 15/42559	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)	(Earliest) Priority Date *(day/month/year)* 30 July 2014 (30.07.2014)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 1

☒ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42559

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42559

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-72, 76-142
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 15/42559

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01)
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01); USPC - 435/91.2; 435/91.1; 435/6.1; 435/6.12; 536/23.1
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01); USPC - 435/91.2; 435/91.1; 435/6.1; 435/6.12; 536/23.1
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00 - see keyword below

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(USPT,PGPB,EPAB,JPAB); PatBase; Medline, Google: oligonucleotide, amplification, multiplex, plurality, library, pool, unique, PCR, real-time PCR, RT-PCR, reverse transcribing, producing, RNA, DNA, cDNA, selective, degrading, alkaline hydrolysis, RNase H, RNase T, DNA polymerase, Taq, RNA polymerase, T7 polymerase, fragment, aptamer, length

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2012/0035065 A1 (SMOLKE et al.) 09 February 2012 (09.02.2012), para [0007], [0032], [0047], [0061], [0083], [0091], [0104], [0117], [0119], [0121], [0122], [0123], [0124], [0125], and [0147]	1-2, 73-74 ---------- 3, 75
Y	US 2013/0096014 A1 (ANDERSEN et al.) 18 April 2013 (18.04.2013), para [0005], [0006], [0010], [0035], [0061], and [0067]	3, 75
A	FAKRUDDIN et al., NUCLEIC ACID SEQUENCE BASED AMPLIFICATION (NASBA)-PROSPECTS AND APPLICATIONS. Intern J Life Sci Pharm Res. 2012, Vol 2(1), p. L106-121. Entire documentation, especially Abstract; and p. L108, Fig 1	1-3, 73-75
A	SHIELS et al., RNA-DNA Hybrids Containing Damaged DNA are Substrates for RNase H. Bioorg Med Chem Lett. 2001, Vol. 11(19), p. 2623-6. Entire documentation, especially Abstract; and pg 2625, col 1, top para	1-3, 73-75

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 September 2015 (10.09.2015)	09 OCT 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	09 OCT 2015

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498.70510	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 15/42559	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01)
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 *bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to th IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Fom PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	10 September 2015 (10.09.2015)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-72, 76-142

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 4-72, 76-142 are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 4-72, 76-142 are not drafted in accordance with the second and third sentences of Rule 6.4 (a). These claims are improper multiple dependent claims.

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-72, 76-142

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-3, 73-75	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-3, 73-75	NO
Industrial applicability (IA)	Claims	1-3, 73-75	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-2 and 73-74 lack an inventive step under PCT Article 33(3) as being obvious over US 2012/0035065 A1 to SMOLKE et al. (hereinafter 'Smolke').

Regarding claim 1, Smolke discloses a method (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR'; para [0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'), including:
--amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ...a common oligonucleotide pool for aptamer ... 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers... oligonucleotide pool is amplified using standard PCR techniques', indicating 'amplifying at least some of a plurality of oligonucleotides'; para [0007] - 'aptamer-regulated nucleic acid sensors... ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'; para [0061] - 'ampliSwitch sensing events are amplified using triggered PCR. ... primer:template pair, and analyte...detected through quantitative real-time PCR ... oligonucleotide'; para [0147] - 'libraries of ampliSwitches... simultaneously...multiplex assays can be run in a single reaction tube'; para [0117] - 'single-stranded target nucleic acid template ...first NASBA primer is extended to form an extension product... DNA'; para [0083] - 'primer:template pair ... a DNA polymerase ... taq polymerase ... extension product'; para [0125]);
--transcribing in vitro at least some of the amplified oligonucleotides to produce RNA (para [0091] - 'in vitro ... identifying RNA or DNA aptamers involve first preparing a large pool of DNA... oligonucleotide pool is amplified using standard PCR ... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired'; para [0104] - 'primer extension may be performed by PCR,... NASBA, transcription-mediated amplification'; para [0125] - 'NASBA reaction requires,... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter'; para [0123] - 'inclusion of an RNA polymerase and the required four ribonucleoside triphosphates ... results in the synthesis of an RNA strand .....DNA-directed RNA polymerase... synthesizes several copies of RNA...T7 RNA polymerase', wherein 'T7 RNA polymerase' transcribing DNA 'oligonucleotides to produce RNA'; Please see Wikipedia_T7_RNA_polymerase - 'T7 RNA Polymerase is an RNA polymerase ...catalyzes the formation of RNA'; Activity - 'T7 polymerase... transcribes only DNA downstream of a T7 promoter'; Reference is not provided as it is a common knowledge);
--reverse transcribing the RNA to produce transcribed DNA (para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase', wherein 'a reverse transcriptase' 'reverse transcribing the RNA to produce transcribed DNA'; para [0091] - 'oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR...The cDNA is then amplified'; Specification: pg 7, ln 13-15 - 'plurality of RNAs 30, may then be used to produce additional amounts of DNA 40, e.g., by using reverse transcription...reverse transcriptase, may be used to perform the reverse transcription'); and
--selectively degrading the RNA relative to the transcribed DNA (para [0125] - 'NASBA reaction requires...a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0119] - 'NASBA reaction also includes an RNA degrading enzyme ... that will hydrolyze the RNA of an RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, ... RNase H', wherein 'hydrolyze the RNA of an RNA:DNA hybrid' is 'selectively degrading the RNA relative to the transcribed DNA').

Although Smolke does not expressively disclose the method comprising about steps in the order as claimed, one of the ordinary skill in the art at the time the invention was made would have known to obtain a method comprising the steps in the order as claimed based on the need, in order to arrange methods available in the art for facilitating generating a plurality of DNA oligonucleotides with an expected success and without undue experimentation, especially Smolke discloses all steps can be used in no particular order with materials that are needed for carrying out each step for different purposes (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ...oligonucleotide pool is amplified... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts ...after reverse transcription.. cDNA is then amplified'; para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0007]; para [0047]; para [0061]; para [0147]; para [0104]; para [0123], and para [0119], see above for all quotations and explanations).

**********Continued in the supplemental box**********

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.

PCT/US 15/42559

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

Note

I) Claim 26 is objected to as it is partially self-referring and partially depending from the following claim 27. It is assumed that claim 26 depends from any one of claims 1-25. For the purposes of this ISR, claim 26 is construed as follows:

26. The method of any one of claims 1-25, wherein transcribing at least some of the amplified oligonucleotides comprises a mass of RNA that is at least 100-fold greater than the mass of amplified oligonucleotides.

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V No 2 (citations and explanations)

Regarding claim 2, Smolke further discloses the method of claim 1, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides (para [0091] - 'in vitro selection techniques for identifying RNA or DNA aptamers involve first preparing a large pool of DNA molecules of the desired length...a common oligonucleotide pool for aptamer selection might contain a region of 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers. The oligonucleotide pool is amplified using standard PCR techniques, ... The DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired...oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR... cDNA is then amplified', wherein '20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers' comprising 'the plurality of oligonucleotides have an length of between 50 [20+15x2] and 150 [100+25x2] nucleotides , which are in the range of 'the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides').

Regarding claim 73, Smolke discloses a method (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR'; para[0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'), including:
--simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ...a common oligonucleotide pool for aptamer ... a region of 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers...oligonucleotide pool is amplified using standard PCR techniques', indicating 'simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides'; para [0007] - 'aptamer-regulated nucleic acid sensors... ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'; para [0147] - 'libraries of ampliSwitches... simultaneously...multiplex assays can be run in a single reaction tube'; para [0117] - 'single-stranded target nucleic acid template ...first NASBA primer is extended to form an extension product... DNA'; para [0083] - 'primer:template pair . a DNA polymerase ... taq polymerase ... extension product'; para [0125]);
--transcribing in vitro at least some of the amplified oligonucleotides to produce RNA (para [0091] - 'in vitro ... identifying RNA or DNA aptamers involve first preparing a large pool of DNA... oligonucleotide pool is amplified using standard PCR ... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired'; para [0104] - 'primer extension may be performed by PCR,... NASBA, transcription-mediated amplification'; para [0125] - 'NASBA reaction requires,... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter'; para [0123] - 'inclusion of an RNA polymerase and the required four ribonucleoside triphosphates ... results in the synthesis of an RNA strand .....DNA-directed RNA polymerase... synthesizes several copies of RNA...T7 RNA polymerase', wherein 'T7 RNA polymerase' transcribing DNA 'oligonucleotides to produce RNA'; Please see Wikipedia_T7_RNA_polymerase - 'T7 RNA Polymerase is an RNA polymerase ...catalyzes the formation of RNA'; Activity - 'T7 polymerase... transcribes only DNA downstream of a T7 promoter'; Reference is not provided as it is a common knowledge);
--reverse transcribing the RNA to produce transcribed DNA (para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase', wherein 'a reverse transcriptase' 'reverse transcribing the RNA to produce transcribed DNA'; para [0091] - 'oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR...The cDNA is then amplified'; Specification: pg 7, ln 13-15 - 'plurality of RNAs 30, may then be used to produce additional amounts of DNA 40, e.g., by using reverse transcription...reverse transcriptase, may be used to perform the reverse transcription'); and
--selectively degrading the RNA relative to the transcribed DNA (para [0125] - 'NASBA reaction requires...a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0119] - 'NASBA reaction also includes an RNA degrading enzyme ... that will hydrolyze the RNA of an RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, ... RNase H', wherein 'hydrolyze the RNA of an RNA:DNA hybrid' is 'selectively degrading the RNA relative to the transcribed DNA').

Although Smolke does not expressively disclose the method comprising about steps in the order as claimed, one of the ordinary skill in the art at the time the invention was made would have known to obtain a method comprising the steps in the order as claimed based on the need, in order to arrange methods available in the art for facilitating generating a plurality of DNA oligonucleotides with an expected success and without undue experimentation, especially Smolke discloses all steps can be used in no particular order with materials that are needed for carrying out each step for different purposes (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ... oligonucleotide pool is amplified... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts ...after reverse transcription.. cDNA is then amplified'; para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0007]; para [0047]; para [0061]; para [0147]; para [0104]; para [0123], and para [0119], see above for all quotations and explanations).

**********Continued in the next supplemental box**********

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
The previous supplemental box - Box V No 2 (citations and explanations)

Regarding claim 74, Smolke further discloses the method of claim 73, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides (para [0091] - 'in vitro selection techniques for identifying RNA or DNA aptamers involve first preparing a large pool of DNA molecules of the desired length...a common oligonucleotide pool for aptamer selection might contain a region of 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers. The oligonucleotide pool is amplified using standard PCR techniques, ... The DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired...oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR... cDNA is then amplified', wherein '20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers' comprising 'the plurality of oligonucleotides have an length of between 50 [20+15x2] and 150 [100+25x2] nucleotides, which are in the range of 'the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides').

Claims 3 and 75 lack an inventive step under PCT Article 33(3) as being obvious over Smolke, as above, in view of US 2013/0096014 A1 to ANDERSEN et al. (hereinafter 'Andersen').

Regarding claim 3, Smolke discloses a method, comprising: amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, as discussed in the discussion of any one of claims 1 or 2. Smolke further discloses wherein the plurality of oligonucleotides includes two or more unique oligonucleotide sequences (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR techniques'; para [0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0032] - 'a library of two or more ampliSwitch nucleic acids, ...comprises (i) a priming sequence that hybridizes to a target template to form a primer:template pair, and (ii) an aptamer that binds to a ligand.... each of the two or more ampliSwitch nucleic acids comprising a unique priming sequence ..each ampliSwitch molecule's priming sequence hybridizes to one or more unique target templates'; para [0047] - 'a set of ampliSwitches for the simultaneous detection of different analytes using ampliSwitch sensors. ... primer:template pair... through qRT-PCR').

Smolke does not specifically teach wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences. Andersen discloses a method for amplifying a plurality of polynucleotide sequences including both DNA and RNA in real time by using a plurality of amplification primer pairs or sets, each of which is for amplifying a different polynucleotide sequence of interest (para [0006] - 'a plurality of different sequences are amplified simultaneously in a single reaction'; para [0005] - 'amplifying polynucleotide sequences ... multiplex .... using a plurality of amplification primer pairs or sets, each of which is suitable or operative for amplifying a different polynucleotide sequence of interest'; para [0010] - 'Both DNA and RNA target polynucleotides can be multiplex amplified using such low primer concentrations.... reverse transcription of RNA into cDNA via a reverse-transcription and subsequent multiplex amplification of the resultant cDNA with a DNA polymerase'; para [0061] - 'a multiplex amplification reaction... can be assessed using real-time PCR methods. ... amplifying a plurality of polynucleotides ... in a multiplex PCR amplification reaction using a plurality of primer sets'), wherein at least 100 different primer pairs can be used by using low concentration of each primer pair, each of which is used for amplifying a different target sequence of interest (para [0035] - 'one or more RNAs or DNAs, is amplified by PCR or RT-PCR in the presence of a plurality of amplification primer pairs, each of which is suitable for amplifying a different target sequence of interest.... the number of primer pairs can be at least 100'; para [0067] - 'highly efficient multiplex amplification was achieved with primer concentrations as low as 45 nM each primer').

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of Smolke and Andersen, to obtain a method, comprising: amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, based on the teaching of Smolke, and further wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences, based on the combination of Andersen and Smolke, in order to combine the technologies available in the art for facilitating producing a plurality of transcribed DNA oligonucleotides in real time with an expected success and without undue experimentation.

**********Continued in the next supplemental box**********

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
The previous supplemental box - Box V No 2 (citations and explanations)

Regarding claim 75, Smolke discloses a method, comprising: simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, as discussed in the discussion of any one of claims 73 or 74. Smolke further discloses wherein the plurality of oligonucleotides includes two or more unique oligonucleotide sequences (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR techniques'; para [0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0032] - 'a library of two or more ampliSwitch nucleic acids ...comprises (i) a priming sequence that hybridizes to a target template to form a primer:template pair, and (ii) an aptamer that binds to a ligand.... each of the two or more ampliSwitch nucleic acids comprising a unique priming sequence ..each ampliSwitch molecule's priming sequence hybridizes to one or more unique target templates'; para [0047] - 'a set of ampliSwitches for the simultaneous detection of different analytes using ampliSwitch sensors. ... primer:template pair... through qRT-PCR').

Smolke does not specifically teach wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences. Andersen discloses a method for amplifying a plurality of polynucleotide sequences including both DNA and RNA in real time by using a plurality of amplification primer pairs or sets, each of which is for amplifying a different polynucleotide sequence of interest (para [0006] - 'a plurality of different sequences are amplified simultaneously in a single reaction'; para [0005] - 'amplifying polynucleotide sequences ... multiplex .... using a plurality of amplification primer pairs or sets, each of which is suitable or operative for amplifying a different polynucleotide sequence of interest'; para [0010] - 'Both DNA and RNA target polynucleotides can be multiplex amplified using such low primer concentrations.... reverse transcription of RNA into cDNA via a reverse-transcription and subsequent multiplex amplification of the resultant cDNA with a DNA polymerase'; para [0061] - 'a multiplex amplification reaction... can be assessed using real-time PCR methods. ... amplifying a plurality of polynucleotides ... in a multiplex PCR amplification reaction using a plurality of primer sets'), wherein at least 100 different primer pairs can be used by using low concentration of each primer pair, each of which is used for amplifying a different target sequence of interest (para [0035] - 'one or more RNAs or DNAs, is amplified by PCR or RT-PCR in the presence of a plurality of amplification primer pairs, each of which is suitable for amplifying a different target sequence of interest.... the number of primer pairs can be at least 100'; para [0067] - 'highly efficient multiplex amplification was achieved with primer concentrations as low as 45 nM each primer').

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of Smolke and Andersen, to obtain a method, comprising: simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, based on the teaching of Smolke, and further wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences, based on the combination of Andersen and Smolke, in order to combine the technologies available in the art for facilitating producing a plurality of transcribed DNA oligonucleotides simultaneously with an expected success and without undue experimentation.

Claims 1-3, 73-75 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

SEARCH HISTORY

Application Number	PCT/US 15/42559
Search Conducted By	LIC
Search Approved By	LIC

IPC/CPC/US Classifications Searched	IPC(8): C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01) CPC: C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00 USPC: 435/91.2; 435/91.1; 435/6.1; 435/6.12; 536/23.1
Date Conducted	10 September 2015

Documentation Searched	IPC(8): C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01) CPC: C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00 USPC: 435/91.2; 435/91.1; 435/6.1; 435/6.12; 536/23.1 (keyword limited; terms below)
Search Terms Used	oligonucleotide, amplification, multiplex, plurality, library, pool, unique, PCR, real-time PCR, RT-PCR, reverse transcribing, producing, RNA, DNA, cDNA, selective, degrading, alkaline hydrolysis, RNase H, RNase T, DNA polymerase, Taq, RNA polymerase, T7 polymerase, fragment, aptamer, length, nucleotide, bp, nucleic acid, 10, 100, 200, in vitro, polynucleotide
Date Conducted	10 September 2015

Electronic Database Searched	PubWEST
Files Searched	PGPB, USPT, EPAB, JPAB
Date Conducted	10 September 2015
Search Logic:	

DATE: Wednesday, September 09, 2015 Purge Queries Printable Copy Create Case

Order By	Set Name Side by Side	Query	Hit Count	Set Name Result Set	Set Name Grid	Set Name Classification

DB=PGPB,USPT,EPAB,JPAB; PLUR=YES; OP=ADJ

SEARCH HISTORY

Date	L19	(L13 or L12) and (T7 polymerase or RNA polymerase)	12	L19	L19	L19
Date	L18	(L13 or L12) and (oligonucleotides aptamer or Multiplex) same (nucleotide or bp or nucleic acid or "10" or "100" or "200" or length)	13	L18	L18	L18
Date	L17	(L13 or L12) and (oligonucleotides or fragment or amplif$ or length) same (nucleotide or bp or nucleic acid or "10" or "100" or "200")	13	L17	L17	L17
Date	L16	(L13 or L12) and L11	12	L16	L16	L16
Date	L15	(L13 or L12) and ((DNA polymerase or Taq)or (produc$ adj20 DNA) or (DNA adj20 produc$))	13	L15	L15	L15
Date	L14	(L13 or L12) and ((degrad$ or alkaline hydrolysis or RNase H or RNase T) adj20 RNA or RNA adj20 (degrad$ or alkaline hydrolysis or RNase H or RNase T))	13	L14	L14	L14
Date	L13	US 20130303407 A1 or US 20130096014 A1 or US 20120142060 A1 or US 20110081685 A1 or US 20080227661 A1 or US 20100021973 A1 or US 20070031857 A1	11	L13	L13	L13
Date	L12	US 20120035065 A1	2	L12	L12	L12
Date	L11	L10 and L1	21	L11	L11	L11
Date	L10	L9 and @pd<20140730	29	L10	L10	L10
Date	L9	L8 and ((degrad$ or alkaline hydrolysis or RNase H or RNase T) adj20 RNA or RNA adj20 (degrad$ or alkaline hydrolysis or RNase H or RNase T))	36	L9	L9	L9
Date	L8	L7 and (reverse transcri$ or rt-PCR or rtPCR) same (DNA or cDNA)	92	L8	L8	L8
Date	L7	L5 and ((amplifying or amplified) adj20 oligonucleotides or oligonucleotides adj20 (amplifying or amplified)) same produc$ same RNA	135	L7	L7	L7
Date	L6	L5 and ((amplifying or amplified) adj20 oligonucleotides or oligonucleotides adj20 (amplifying or amplified))	1260	L6	L6	L6

SEARCH HISTORY

Date	L5	L4 same oligonucleotides	2378	L5	L5	L5
Date	L4	Multiplex adj20 (PCR or real-time PCR or rt-PCR or rtPCR) or (PCR or real-time PCR or rt-PCR or rtPCR) adj20 multiplex	10471	L4	L4	L4
Date	L3	(Multiplex or library) adj20 (PCR or real-time PCR or rt-PCR or rtPCR) or (PCR or real-time PCR or rt-PCR or rtPCR) adj20 (multiplex or library)	56573	L3	L3	L3
Date	L2	(435/91.2 or 435/91.1 or 435/6.1 or 435/6.12 or 536/23.1)[CCLS]	57200	L2	L2	L2
Date	L1	(435/91.2 or 435/91.1 or 435/6.1 or 435/6.12 or 536/23.1)![CCLS]	57200	L1	L1	L1

END OF SEARCH HISTORY

Search 2:

DATE: Wednesday, September 09, 2015 Purge Queries Printable Copy Create Case

Order By	Set Name Side by Side	Query	Hit Count	Set Name Result Set	Set Name Grid	Set Name Classification
		DB=PGPB,USPT,EPAB,JPAB; PLUR=YES; OP=ADJ				
Date	L16	L1 and template same (first or second or third)	5	L16	L16	L16
Date	L15	L1 and NASBA	4	L15	L15	L15
Date	L14	L1 and (transcribing or in vitro or RNA or T7 polymerase or RNA polymerase)	5	L14	L14	L14
Date	L13	L1 and (transcribing or in vitro RNA or T7 polymerase or RNA polymerase)	4	L13	L13	L13
Date	L12	L1 and (real-time or real time) same PCR	5	L12	L12	L12
Date	L11	L1 and (T7 polymerase or RNA polymerase)	4	L11	L11	L11
Date	L10	L1 and (T7 or RNA) adj20 polymerase	5	L10	L10	L10
Date	L9	L1 and((degrad$ or alkaline hydrolysis or RNase H or RNase T) adj20 RNA or RNA adj20 (degrad$ or alkaline hydrolysis or RNase H or RNase T))	5	L9	L9	L9
Date	L8	L1 and (DNA polymerase or Taq)	5	L8	L8	L8

Date	L7	L1 and oligonucleotide same (amplif$ or PCR or RT-PCR)	5	L7	L7	L7
Date	L6	L1 and length	4	L6	L6	L6
Date	L5	L1 and (simultaneous$ or plurality) same (amplif$ or PCR or RT-PCR)	5	L5	L5	L5
Date	L4	L1 and simultaneous$	5	L4	L4	L4
Date	L3	L1 and (produc$ or amplif$) same (DNA or RNA)	5	L3	L3	L3
Date	L2	L1 and Multiplex same (oligonucleotides or aptamer or amplif$ or PCR)	5	L2	L2	L2
Date	L1	US 20120035065 A1 or US 20130096014 A1 or US 20100021973 A1	5	L1	L1	L1

END OF SEARCH HISTORY

Search 3:

DATE: Thursday, September 10, 2015 Purge Queries Printable Copy Create Case

Order By	Set Name Side by Side	Query	Hit Count	Set Name Result Set	Set Name Grid	Set Name Classification
		DB=PGPB,USPT,EPAB,JPAB; PLUR=YES; OP=ADJ				
Date	L12	L1 and oligonucleotide same (polynucleotide or nucleotide)	3	L12	L12	L12
Date	L11	L1 and oligonucleotide	3	L11	L11	L11
Date	L10	L9 and (simultaneous$ or multiplex) same amplif$ same (DNA or RNA)	1	L10	L10	L10
Date	L9	L1 and (simultaneous$ or multiplex) same (amplif$ or PCR or "10" or "100")	3	L9	L9	L9
Date	L8	L6 and multiplex same (amplif$ or PCR or "10" or "100")	2	L8	L8	L8
Date	L7	L6 and oligonucleotide same (library or pool or unique)	2	L7	L7	L7
Date	L6	L1 and NASBA same (oligonucleotide or aptamer)	2	L6	L6	L6
Date	L5	L1 and (ampliSwitch or aptamer) same (library or detect$ or target)	2	L5	L5	L5
Date	L4	L1 and method same (amplif$ or PCR)	3	L4	L4	L4

SEARCH HISTORY

Date	L3	L1 and unique same (target or nucleotide or sequence or two or more or ten or "10")	2	L3	L3	L3
Date	L2	L1 and (reverse transcri$ or rt-PCR or rtPCR) same (DNA or cDNA)	3	L2	L2	L2
Date	L1	US 20120035065 A1 or US 20130096014 A1	3	L1	L1	L1

END OF SEARCH HISTORY

Electronic Database Searched	PatBase
Files Searched	Full-text: AU BE BR CA CH CN DE DK EP ES FI FR GB IN JP KR SE TH TW US WO Bibliographic: (Europe) AT BA BE BG CH CS CY CZ DD DK EE ES FI GE GR HR HU IE IS IT LT LU LV MC MD MT NL NO PL PT RO RS SE SI SK SM TR UA YU (Asia) EA GC HK ID IL IN KZ MN MY PH RU SG SU TH TJ TW UZ VN (North America) CA CR CU DO GT HN MX NI PA SV TT (South America) AR BR CL CO EC PE UY (Australasia) AU NZ (Africa) AP DZ EG KE MA MW OA ZA ZM ZW
Date Conducted	09 September 2015
Search Logic:	

1) IC=(C12P19/34 or C12Q1/68 or C07H21/04) (100000)
2) CPC=(C12Q1/686 or C12N15/10 or C12Q1/6844 or C07H21/00) (9384)
3) FT=(Multiplex w20 (PCR or real-time PCR or rt-PCR or rtPCR) or library w20 (PCR or real-time PCR or rt-PCR or rtPCR)) (7793)
4) 3 and ft=(Multiplex w20 (PCR or real-time PCR or rt-PCR or rtPCR)) (7793)
5) 4 and ft=(oligonucleotides w20 (PCR or real-time PCR or rt-PCR or rtPCR)) (2624)
6) 5 and ft=((amplifying or amplified) w20 oligonucleotides) (1249)
7) 6 and ft=(produc* w20 RNA) (553)
8) 7 and ft=(((reverse transcri$ or rt-PCR or rtPCR) w20 DNA) or ((reverse transcri$ or rt-PCR or rtPCR) w20 cDNA)) (190)
9) 8 and ft=((degrad* or alkaline hydrolysis or RNase H or RNase T) w20 RNA) (103)
10) 9 and pd<20140730 (101)
11) 10 and ft=(DNA polymerase or Taq) (96)
12) 11 and ft=(T7 polymerase or RNA polymerase) (82)
13) 12 and (1 or 2) (73)
14) 13 and 2 (16)

SEARCH HISTORY

Electronic Database Searched	Google
Files Searched	Google Scholar, Current Contents, Dissertation Abstracts, Inside Conferences (Non-Patent Literature), WIPO
Date Conducted	10 September 2015

Search Logic:

1) amplify oligonucleotide producing RNA: About 13,300,000 results
2) transcribing oligonucleotide producing RNA: About 15,100,000 results
3) in vitro transcription reverse transcription: About 27,900,000 results
4) real-time PCR reverse transcription selective degrading: About 666,000 results
5) real-time PCR amplifying oligonucleotide simultaneously two or more: About 638,000 results
6) real-time PCR amplifying oligonucleotide selectively degrading RNA: About 644,000 results
7) oligonucleotide real-time PCR amplification Multiplex: About 319,000 results
8) pcr dna amplification DNA polymerase: About 2,600,000 results
9) pcr dna amplification T7 polymerase RNA production: About 458,000 results
10) nasba amplification: About 183,000 results
11) multiplex PCR polymerase RNA DNA reverse transcription: About 385,000 results
12) RNA degrading RNAse H alkaline hydrolysis: About 235,000 results

Electronic Database Searched	PubMed, NCBI
Files Searched	Non-Patent Literature, Database
Date Conducted	10 September 2015

Search Logic:

- reverse transcription rna cdna AND (amplification OR amplifying) ... (11)

  reverse transcription rna cdna AND (amplification OR amplifying) AND (degrading OR degradation) AND oligonucleotide AND (real time OR real-time OR rt-PCR OR rtPCT) AND mixture

SEARCH HISTORY

PubMed

- reverse transcription rna cdna AND (amplification OR amplifying) ... (0)

  reverse transcription rna cdna AND (amplification OR amplifying) AND (degrading OR degradation) AND oligonucleotide AND (real time OR real-time OR rt-PCR OR rtPCT) AND plurality
  PubMed
- reverse transcription rna cdna AND (amplification OR amplifying) ... (222)

  reverse transcription rna cdna AND (amplification OR amplifying) AND (degrading OR degradation) AND oligonucleotide AND (real time OR real-time OR rt-PCR OR rtPCT)
  PubMed
- reverse transcription rna cdna AND (amplification OR amplifying) ... (341)

  reverse transcription rna cdna AND (amplification OR amplifying) AND (degrading OR degradation) AND oligonucleotide
  PubMed
- reverse transcription rna cdna AND (amplification OR amplifying) ... (1131)

  reverse transcription rna cdna AND (amplification OR amplifying) AND (degrading OR degradation)
  PubMed

PATENT COOPERATION TREATY

ADVANCE E-MAIL

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING TRANSMITTAL OF COPY OF INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY (CHAPTER I OF THE PATENT COOPERATION TREATY)
(PCT Rule 44bis.1(c))

To:

CHEN, Tani
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)* 09 February 2017 (09.02.2017)	
Applicant's or agent's file reference H0498.70510	**IMPORTANT NOTICE**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2015/042559	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

Applicant
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Mineko Mohri

e-mail: pct.team8@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference H0498.70510	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2015/042559	International filing date *(day/month/year)* 29 July 2015 (29.07.2015)	Priority date *(day/month/year)* 30 July 2014 (30.07.2014)
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 9 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☒	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☒	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 31 January 2017 (31.01.2017)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Mineko Mohri e-mail: pct.team8@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	09 OCT 2015

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498.70510	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 15/42559	29 July 2015 (29.07.2015)	30 July 2014 (30.07.2014)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - C12P 19/34; C12Q 1/68; C07H 21/04 (2015.01)
CPC - C12Q 1/686; C12N 15/10; C12Q 1/6844; C07H 21/00

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☒ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1 *bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to th IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Fom PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	10 September 2015 (10.09.2015)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [x] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. [x] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:
   - a. [x] forming part of the international application as filed:
     - [x] in the form of an Annex C/ST.25 text file.
     - [ ] on paper or in the form of an image file.
   - b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   - c. [ ] furnished subsequent to the international filing date for the purposes of international search only:
     - [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
     - [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 4-72, 76-142

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 4-72, 76-142 are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 4-72, 76-142 are not drafted in accordance with the second and third sentences of Rule 6.4 (a). These claims are improper multiple dependent claims.

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 4-72, 76-142

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-3, 73-75	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-3, 73-75	NO
Industrial applicability (IA)	Claims	1-3, 73-75	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-2 and 73-74 lack an inventive step under PCT Article 33(3) as being obvious over US 2012/0035065 A1 to SMOLKE et al. (hereinafter 'Smolke').

Regarding claim 1, Smolke discloses a method (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR'; para [0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'), including:
--amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ...a common oligonucleotide pool for aptamer ... 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers... oligonucleotide pool is amplified using standard PCR techniques', indicating 'amplifying at least some of a plurality of oligonucleotides'; para [0007] - 'aptamer-regulated nucleic acid sensors... ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'; para [0061] - 'ampliSwitch sensing events are amplified using triggered PCR. ... primer:template pair, and analyte...detected through quantitative real-time PCR ... oligonucleotide'; para [0147] - 'libraries of ampliSwitches... simultaneously...multiplex assays can be run in a single reaction tube'; para [0117] - 'single-stranded target nucleic acid template ...first NASBA primer is extended to form an extension product... DNA'; para [0083] - 'primer:template pair ... a DNA polymerase ... taq polymerase ... extension product'; para [0125]);
--transcribing in vitro at least some of the amplified oligonucleotides to produce RNA (para [0091] - 'in vitro ... identifying RNA or DNA aptamers involve first preparing a large pool of DNA... oligonucleotide pool is amplified using standard PCR ... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired'; para [0104] - 'primer extension may be performed by PCR,... NASBA, transcription-mediated amplification'; para [0125] - 'NASBA reaction requires,... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter'; para [0123] - 'inclusion of an RNA polymerase and the required four ribonucleoside triphosphates ... results in the synthesis of an RNA strand .....DNA-directed RNA polymerase... synthesizes several copies of RNA...T7 RNA polymerase', wherein 'T7 RNA polymerase' transcribing DNA 'oligonucleotides to produce RNA'; Please see Wikipedia_T7_RNA_polymerase - 'T7 RNA Polymerase is an RNA polymerase ...catalyzes the formation of RNA'; Activity - 'T7 polymerase... transcribes only DNA downstream of a T7 promoter'; Reference is not provided as it is a common knowledge);
--reverse transcribing the RNA to produce transcribed DNA (para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase', wherein 'a reverse transcriptase' 'reverse transcribing the RNA to produce transcribed DNA'; para [0091] - 'oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR...The cDNA is then amplified'; Specification: pg 7, ln 13-15 - 'plurality of RNAs 30, may then be used to produce additional amounts of DNA 40, e.g., by using reverse transcription...reverse transcriptase, may be used to perform the reverse transcription'); and
--selectively degrading the RNA relative to the transcribed DNA (para [0125] - 'NASBA reaction requires...a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0119] - 'NASBA reaction also includes an RNA degrading enzyme ... that will hydrolyze the RNA of an RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, ... RNase H', wherein 'hydrolyze the RNA of an RNA:DNA hybrid' is 'selectively degrading the RNA relative to the transcribed DNA').

Although Smolke does not expressively disclose the method comprising about steps in the order as claimed, one of the ordinary skill in the art at the time the invention was made would have known to obtain a method comprising the steps in the order as claimed based on the need, in order to arrange methods available in the art for facilitating generating a plurality of DNA oligonucleotides with an expected success and without undue experimentation, especially Smolke discloses all steps can be used in no particular order with materials that are needed for carrying out each step for different purposes (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ...oligonucleotide pool is amplified... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts ...after reverse transcription.. cDNA is then amplified'; para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0007]; para [0047]; para [0061]; para [0147]; para [0104]; para [0123], and para [0119], see above for all quotations and explanations).

**********Continued in the supplemental box**********

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Box No. VIII Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

Note

I) Claim 26 is objected to as it is partially self-referring and partially depending from the following claim 27. It is assumed that claim 26 depends from any one of claims 1-25. For the purposes of this ISR, claim 26 is construed as follows:

26. The method of any one of claims 1-25, wherein transcribing at least some of the amplified oligonucleotides comprises a mass of RNA that is at least 100-fold greater than the mass of amplified oligonucleotides.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V No 2 (citations and explanations)

Regarding claim 2, Smolke further discloses the method of claim 1, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides (para [0091] - 'in vitro selection techniques for identifying RNA or DNA aptamers involve first preparing a large pool of DNA molecules of the desired length...a common oligonucleotide pool for aptamer selection might contain a region of 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers. The oligonucleotide pool is amplified using standard PCR techniques, ... The DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired...oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR... cDNA is then amplified', wherein '20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers' comprising 'the plurality of oligonucleotides have an length of between 50 [20+15x2] and 150 [100+25x2] nucleotides , which are in the range of 'the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides').

Regarding claim 73, Smolke discloses a method (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR'; para[0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'), including:
--simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ...a common oligonucleotide pool for aptamer ... a region of 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers...oligonucleotide pool is amplified using standard PCR techniques', indicating 'simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides'; para [0007] - 'aptamer-regulated nucleic acid sensors... ampliSwitches'; para [0047] - 'set of ampliSwitches ... simultaneous detection of different analytes ... primer:template pair... qRT-PCR'; para [0147] - 'libraries of ampliSwitches... simultaneously...multiplex assays can be run in a single reaction tube'; para [0117] - 'single-stranded target nucleic acid template ...first NASBA primer is extended to form an extension product... DNA'; para [0083] - 'primer:template pair . a DNA polymerase ... taq polymerase ... extension product'; para [0125]);
--transcribing in vitro at least some of the amplified oligonucleotides to produce RNA (para [0091] - 'in vitro ... identifying RNA or DNA aptamers involve first preparing a large pool of DNA... oligonucleotide pool is amplified using standard PCR ... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired'; para [0104] - 'primer extension may be performed by PCR,... NASBA, transcription-mediated amplification'; para [0125] - 'NASBA reaction requires,... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter'; para [0123] - 'inclusion of an RNA polymerase and the required four ribonucleoside triphosphates ... results in the synthesis of an RNA strand .....DNA-directed RNA polymerase... synthesizes several copies of RNA...T7 RNA polymerase', wherein 'T7 RNA polymerase' transcribing DNA 'oligonucleotides to produce RNA'; Please see Wikipedia_T7_RNA_polymerase - 'T7 RNA Polymerase is an RNA polymerase ...catalyzes the formation of RNA'; Activity - 'T7 polymerase... transcribes only DNA downstream of a T7 promoter'; Reference is not provided as it is a common knowledge);
--reverse transcribing the RNA to produce transcribed DNA (para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase', wherein 'a reverse transcriptase' 'reverse transcribing the RNA to produce transcribed DNA'; para [0091] - 'oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR...The cDNA is then amplified'; Specification: pg 7, ln 13-15 - 'plurality of RNAs 30, may then be used to produce additional amounts of DNA 40, e.g., by using reverse transcription...reverse transcriptase, may be used to perform the reverse transcription'); and
--selectively degrading the RNA relative to the transcribed DNA (para [0125] - 'NASBA reaction requires...a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0119] - 'NASBA reaction also includes an RNA degrading enzyme ... that will hydrolyze the RNA of an RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, ... RNase H', wherein 'hydrolyze the RNA of an RNA:DNA hybrid' is 'selectively degrading the RNA relative to the transcribed DNA').
Although Smolke does not expressively disclose the method comprising about steps in the order as claimed, one of the ordinary skill in the art at the time the invention was made would have known to obtain a method comprising the steps in the order as claimed based on the need, in order to arrange methods available in the art for facilitating generating a plurality of DNA oligonucleotides with an expected success and without undue experimentation, especially Smolke discloses all steps can be used in no particular order with materials that are needed for carrying out each step for different purposes (para [0091] - 'identifying RNA or DNA aptamers...preparing a large pool of DNA ... oligonucleotide pool is amplified... DNA pool may be in vitro transcribed to produce a pool of RNA transcripts ...after reverse transcription.. cDNA is then amplified'; para [0125] - 'NASBA reaction requires, in no particular order, a first NASBA primer ... a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, ... an RNA degrading enzyme'; para [0007]; para [0047]; para [0061]; para [0147]; para [0104]; para [0123], and para [0119], see above for all quotations and explanations).

**********Continued in the next supplemental box**********

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
The previous supplemental box - Box V No 2 (citations and explanations)

Regarding claim 74, Smolke further discloses the method of claim 73, wherein the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides (para [0091] - 'in vitro selection techniques for identifying RNA or DNA aptamers involve first preparing a large pool of DNA molecules of the desired length...a common oligonucleotide pool for aptamer selection might contain a region of 20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers. The oligonucleotide pool is amplified using standard PCR techniques, ... The DNA pool may be in vitro transcribed to produce a pool of RNA transcripts if an RNA aptamer is desired...oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were generated) again by PCR... cDNA is then amplified', wherein '20-100 randomized nucleotides flanked on both ends by an about 15-25 nucleotide long region of defined sequence useful for the binding of PCR primers' comprising 'the plurality of oligonucleotides have an length of between 50 [20+15x2] and 150 [100+25x2] nucleotides, which are in the range of 'the plurality of oligonucleotides have an average length of between 10 and 200 nucleotides').

Claims 3 and 75 lack an inventive step under PCT Article 33(3) as being obvious over Smolke, as above, in view of US 2013/0096014 A1 to ANDERSEN et al. (hereinafter 'Andersen').

Regarding claim 3, Smolke discloses a method, comprising: amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, as discussed in the discussion of any one of claims 1 or 2. Smolke further discloses wherein the plurality of oligonucleotides includes two or more unique oligonucleotide sequences (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR techniques'; para [0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0032] - 'a library of two or more ampliSwitch nucleic acids, ...comprises (i) a priming sequence that hybridizes to a target template to form a primer:template pair, and (ii) an aptamer that binds to a ligand.... each of the two or more ampliSwitch nucleic acids comprising a unique priming sequence ..each ampliSwitch molecule's priming sequence hybridizes to one or more unique target templates'; para [0047] - 'a set of ampliSwitches for the simultaneous detection of different analytes using ampliSwitch sensors. ... primer:template pair... through qRT-PCR').

Smolke does not specifically teach wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences. Andersen discloses a method for amplifying a plurality of polynucleotide sequences including both DNA and RNA in real time by using a plurality of amplification primer pairs or sets, each of which is for amplifying a different polynucleotide sequence of interest (para [0006] - 'a plurality of different sequences are amplified simultaneously in a single reaction'; para [0005] - 'amplifying polynucleotide sequences ... multiplex .... using a plurality of amplification primer pairs or sets, each of which is suitable or operative for amplifying a different polynucleotide sequence of interest'; para [0010] - 'Both DNA and RNA target polynucleotides can be multiplex amplified using such low primer concentrations.... reverse transcription of RNA into cDNA via a reverse-transcription and subsequent multiplex amplification of the resultant cDNA with a DNA polymerase'; para [0061] - 'a multiplex amplification reaction... can be assessed using real-time PCR methods. ... amplifying a plurality of polynucleotides ... in a multiplex PCR amplification reaction using a plurality of primer sets'), wherein at least 100 different primer pairs can be used by using low concentration of each primer pair, each of which is used for amplifying a different target sequence of interest (para [0035] - 'one or more RNAs or DNAs, is amplified by PCR or RT-PCR in the presence of a plurality of amplification primer pairs, each of which is suitable for amplifying a different target sequence of interest.... the number of primer pairs can be at least 100'; para [0067] - 'highly efficient multiplex amplification was achieved with primer concentrations as low as 45 nM each primer').

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of Smolke and Andersen, to obtain a method, comprising: amplifying at least some of a plurality of oligonucleotides using real-time PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, based on the teaching of Smolke, and further wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences, based on the combination of Andersen and Smolke, in order to combine the technologies available in the art for facilitating producing a plurality of transcribed DNA oligonucleotides in real time with an expected success and without undue experimentation.

**********Continued in the next supplemental box**********

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 15/42559

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
The previous supplemental box - Box V No 2 (citations and explanations)

Regarding claim 75, Smolke discloses a method, comprising: simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, as discussed in the discussion of any one of claims 73 or 74. Smolke further discloses wherein the plurality of oligonucleotides includes two or more unique oligonucleotide sequences (para [0091] - 'identifying RNA or DNA aptamers ...oligonucleotide pool is amplified using standard PCR techniques'; para [0007] - 'aptamer-regulated nucleic acid sensors...ampliSwitches'; para [0032] - 'a library of two or more ampliSwitch nucleic acids ...comprises (i) a priming sequence that hybridizes to a target template to form a primer:template pair, and (ii) an aptamer that binds to a ligand.... each of the two or more ampliSwitch nucleic acids comprising a unique priming sequence ..each ampliSwitch molecule's priming sequence hybridizes to one or more unique target templates'; para [0047] - 'a set of ampliSwitches for the simultaneous detection of different analytes using ampliSwitch sensors. ... primer:template pair... through qRT-PCR').

Smolke does not specifically teach wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences. Andersen discloses a method for amplifying a plurality of polynucleotide sequences including both DNA and RNA in real time by using a plurality of amplification primer pairs or sets, each of which is for amplifying a different polynucleotide sequence of interest (para [0006] - 'a plurality of different sequences are amplified simultaneously in a single reaction'; para [0005] - 'amplifying polynucleotide sequences ... multiplex .... using a plurality of amplification primer pairs or sets, each of which is suitable or operative for amplifying a different polynucleotide sequence of interest'; para [0010] - 'Both DNA and RNA target polynucleotides can be multiplex amplified using such low primer concentrations.... reverse transcription of RNA into cDNA via a reverse-transcription and subsequent multiplex amplification of the resultant cDNA with a DNA polymerase'; para [0061] - 'a multiplex amplification reaction... can be assessed using real-time PCR methods. ... amplifying a plurality of polynucleotides ... in a multiplex PCR amplification reaction using a plurality of primer sets'), wherein at least 100 different primer pairs can be used by using low concentration of each primer pair, each of which is used for amplifying a different target sequence of interest (para [0035] - 'one or more RNAs or DNAs, is amplified by PCR or RT-PCR in the presence of a plurality of amplification primer pairs, each of which is suitable for amplifying a different target sequence of interest.... the number of primer pairs can be at least 100'; para [0067] - 'highly efficient multiplex amplification was achieved with primer concentrations as low as 45 nM each primer').

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of Smolke and Andersen, to obtain a method, comprising: simultaneously amplifying at least some of a plurality of oligonucleotides in a common solution using PCR to produce amplified oligonucleotides; transcribing in vitro at least some of the amplified oligonucleotides to produce RNA; reverse transcribing the RNA to produce transcribed DNA; and selectively degrading the RNA relative to the transcribed DNA, based on the teaching of Smolke, and further wherein the plurality of oligonucleotides includes at least 10 unique oligonucleotide sequences, based on the combination of Andersen and Smolke, in order to combine the technologies available in the art for facilitating producing a plurality of transcribed DNA oligonucleotides simultaneously with an expected success and without undue experimentation.

Claims 1-3, 73-75 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

Electronic Acknowledgement Receipt

EFS ID:	28903593
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/Amanda Figueiredo
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	12-APR-2017
Filing Date:	
Time Stamp:	12:39:22
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.pdf	24819 a9bc361fa12672e18e90971637ebbe239d0c45aa	no	1
Warnings:					

Information:					
2		H049870509US02-IDS-TC.pdf	40079 a244ea7c0aed3a406e4cc0fb0f624977a26bf65a	yes	5
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Transmittal Letter		1	3	
	Information Disclosure Statement (IDS) Form (SB08)		4	5	
Warnings:					
Information:					
3	Other Reference-Patent/App/Search documents	H049870509WO00_Inv_to_Pay_mailed_10-06-2015.pdf	186647 bbf1c63643a4f0f75cc98bfa5239205a7d896f29	no	3
Warnings:					
Information:					
4	Other Reference-Patent/App/Search documents	H049870509WO00_ISR_WO_mailed_12-18-2015.pdf	2618027 7399ff9c17f83a77bbbd98dc09dd0495787abf49	no	19
Warnings:					
Information:					
5	Other Reference-Patent/App/Search documents	H049870509WO00_IPRP_mailed_02-09-2017.pdf	561090 e5584e0736ed10bd9800dd8887afca38c37577c6	no	11
Warnings:					
Information:					
6	Other Reference-Patent/App/Search documents	H049870510WO00_ISR_WO_mailed_10-09-2015.pdf	733805 cf0ba70946b8e3cfd13c08cd0c54fddeb67226c6	no	20
Warnings:					
Information:					
7	Other Reference-Patent/App/Search documents	H049870510WO00_IPRP_mailed_02-09-2017.pdf	550965 628fd78f5a9e44ed2f2bbada76c860f18d8b7f03	no	10
Warnings:					
Information:					

8	Non Patent Literature	Chen_et_al_Science_ePub_2015_v348_p1.pdf	6207451 cfbfa607d1f36521e5a61f9f999a029388b035d9	no	21
Warnings:					
Information:					
9	Non Patent Literature	Fakruddin_et_al_Life_Science_2012_v2_pL106.pdf	990518 5ab17dddb6b9da92d32fefc6cc6cac1e936c893a	no	16
Warnings:					
Information:					
10	Non Patent Literature	Lubeck_et_al_Nat_Meth_2014_v11_p360.pdf	387897 cd41b733b5b5ea157d084bb5f39e8c45ed43e40e	no	2
Warnings:					
Information:					
11	Non Patent Literature	Moffitt_et_al_PNAS_2016_v113_p14456.pdf	1087571 45e6677c6638989a0779d256656367ef8b26bfc7	no	6
Warnings:					
Information:					
12	Non Patent Literature	Moffitt_et_al_PNAS_2016_v113_p11046.pdf	1354123 fc3249bb3ca6ef1936403ac625534362f0485e03	no	6
Warnings:					
Information:					
13	Non Patent Literature	Moffitt_et_al_Meth_Enzymol_2016_v572_p1.pdf	1820366 e0528899e7d1aff32ea79780f71ae510bdccb6ad	no	49
Warnings:					
Information:					
14	Non Patent Literature	Shiels_et_al_Bioorg_Med_Chem_Lett_2001_v11_p2623.pdf	1025766 de9903fa4c3386726f1af0f362592dda6637241d	no	16
Warnings:					
Information:					

15	Non Patent Literature	Lubeck_et_al_Nat_Meth_2014_v11_p360_Supplementary_Info.pdf	2389370 a2ac0a3cd237b678f797a457825163eb6c409ce4	no	14
Warnings:					
Information:					
			Total Files Size (in bytes):	19978494	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [ ] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Form PTO-1449
  Copies of Cited References

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	April 12, 2017	Reg. No.	52,728

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: April 12, 2017 Electronic Signature for Amanda K. Figueiredo: _____/Amanda K. Figueiredo/_____

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	Not Yet Assigned
Art Unit:	Not Yet Assigned

CERTIFICATE OF ELECTRONIC FILING UNDER 37 C.F.R. § 1.8

The undersigned hereby certifies that this paper, along with any paper referred to as being attached or enclosed, is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4), on the 12th day of April, 2017.

_____/Amanda K. Figueiredo/_____
Amanda K. Figueiredo

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed within three months of the date of the entry of the national stage, as set forth in 37 C.F.R. § 1.491, in an international application.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of

presentation of the references should not be construed as an indication of the importance of the references.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
* 15/329,651	01-27-2017	Zhuang et al.	H0498.70510US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
PCT/US2015/042556	**10-06-2015**	**Invitation to Pay Additional Fees**	**H0498.70509WO00**
PCT/US2015/042556	**12-18-2015**	**International Search Report and Written Opinion**	**H0498.70509WO00**
PCT/US2015/042556	**02-09-2017**	**International Preliminary Report on Patentability**	**H0498.70509WO00**
PCT/US2015/042559	**10-09-2015**	**International Search Report and Written Opinion**	**H0498.70510WO00**
PCT/US2015/042559	**02-09-2017**	**International Preliminary Report on Patentability**	**H0498.70510WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

Respectfully submitted,

By: /Tani Chen/

Tani Chen, Sc.D., Reg. No. 52,728
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: H0498.70509US02
Date: April 12, 2017
xx

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 15/329,683

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	140		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	60		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	360		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	187 minus 20 =	* 167	x 40 =	6680	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	11 minus 3 =	* 8	x 210 =	1680			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			400			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				390			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	9710		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)	OR	RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/329,683	01/27/2017		960	H0498.70509US02	176	11

CONFIRMATION NO. 7428

86110
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

FILING RECEIPT

OC000000090809047

Date Mailed: 04/24/2017

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Xiaowei Zhuang, Lexington, MA;
Kok-Hao Chen, Singapore, SINGAPORE;
Alistair Boettiger, Cambridge, MA;
Jeffrey R. Moffitt, Somerville, MA;
Siyuan Wang, Cambridge, MA;

Applicant(s)

President and Fellows of Harvard College, Cambridge, MA;

Assignment For Published Patent Application

President and Fellows of Harvard College, Cambridge, MA

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a 371 of PCT/US15/42556 07/29/2015
which claims benefit of 62/031,062 07/30/2014
and claims benefit of 62/050,636 09/15/2014
and claims benefit of 62/142,653 04/03/2015

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: No

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 04/20/2017
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/329,683**
Projected Publication Date: 08/03/2017
Non-Publication Request: No
Early Publication Request: No
**** SMALL ENTITY ****
Title

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

Preliminary Class

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

U.S. APPLICATION NUMBER NO.	FIRST NAMED INVENTOR	ATTY. DOCKET NO.
15/329,683	Xiaowei Zhuang	H0498.70509US02

86110
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

INTERNATIONAL APPLICATION NO.	
PCT/US2015/042556	
I.A. FILING DATE	PRIORITY DATE
07/29/2015	07/30/2014

CONFIRMATION NO. 7428
371 FORMALITIES LETTER

OC000000090809048

Date Mailed: 04/24/2017

Notification of Insufficiency Under 37 CFR 1.492 and/or 1.495 (DO/EO/US)

The following items have been submitted by the applicant or the International Bureau to the United States Patent and Trademark Office as a Designated / Elected Office (37 CFR 1.495).

- Indication of Small Entity Status
- Priority Document
- Copy of the International Application filed on 01/27/2017
- Copy of the International Search Report filed on 01/27/2017
- Preliminary Amendments filed on 01/27/2017
- Information Disclosure Statements filed on 04/12/2017
- Sequence Listing txt file (EFS or physical medium) filed on 01/27/2017
- U.S. Basic National Fees filed on 01/27/2017
- Priority Documents filed on 01/27/2017
- Specification filed on 01/27/2017
- Claims filed on 01/27/2017
- Abstracts filed on 01/27/2017
- Drawings filed on 01/27/2017
- Application Data Sheet (37 CFR 1.76) filed on 01/27/2017

The following items **MUST** be furnished within the period set forth below:

- Additional claim fees of $**8750** as a small entity, including any required multiple dependent claim fee, are required. Applicant must submit the additional claim fees or cancel the additional claims for which fees are due.
- To avoid abandonment, a surcharge (for late submission of search fee, examination fee or inventor's oath or declaration) as set forth in 37 CFR 1.492(h) of $70 for a small entity in compliance with 37 CFR 1.27 , must be submitted with the missing items identified in this letter.

SUMMARY OF FEES DUE:

Total additional fees required for this application is $**8820** for a Small Entity:

- $**70** Surcharge.

Applicant is notified that the above-identified application contains the deficiencies noted below. No period for reply is set forth in this notice for correction of these deficiencies. However, if a deficiency relates to the inventor's oath or declaration, the applicant must file an oath or declaration in compliance with 37 CFR 1.63, or a substitute statement in compliance with 37 CFR 1.64, executed by or with respect to each actual inventor no later than the expiration of the time period set in the "Notice of Allowability" to avoid abandonment. See 37 CFR 1.495(c).

FORM PCT/DO/EO/923 (371 Formalities Notice)

- Properly executed inventor's oath or declaration for the following inventor(s) has not been submitted: **Xiaowei Zhuang**, **Kok-Hao Chen**, **Alistair Boettiger**, **Jeffrey R. Moffitt**, and **Siyuan Wang**

Total additional claim fee(s) for this application is $**8750**
- $**1680** for **8** independent claims over 3.
- $**6680** for **167** total claims over 20.
- $**390** for multiple dependent claim surcharge.

ALL OF THE ITEMS SET FORTH ABOVE MUST BE SUBMITTED WITHIN TWO (2) MONTHS FROM THE DATE OF THIS NOTICE. FAILURE TO PROPERLY RESPOND WILL RESULT IN ABANDONMENT.

The time period set above may be extended by filing a petition and fee for extension of time under the provisions of 37 CFR 1.136(a).

Applicant is reminded that any communications to the United States Patent and Trademark Office must be mailed to the address given in the heading and include the U.S. application no. shown above (37 CFR 1.5)

Registered users of EFS-Web may alternatively submit their reply to this notice via EFS-Web.
https://sportal.uspto.gov/authenticate/AuthenticateUserLocalEPF.html

For more information about EFS-Web please call the USPTO Electronic Business Center at **1-866-217-9197** or visit our website at http://www.uspto.gov/ebc.

If you are not using EFS-Web to submit your reply, you must include a copy of this notice.

BRIAN A HARDEN

Telephone: (703) 756-1815

FORM PCT/DO/EO/923 (371 Formalities Notice)

MULTIPLE DEPENDENT CLAIM FEE CALCULATION SHEET

Substitute for Form PTO-1360
(For use with Form PTO/SB/06)

Application Number	Filing Date
15329683	

Applicant(s) Xiaowei Zhuang

* May be used for additional claims or amendments

CLAIMS	AS FILED		AFTER FIRST AMENDMENT		AFTER SECOND AMENDMENT			*		*		*	
	Indep	Depend	Indep	Depend	Indep	Depend		Indep	Depend	Indep	Depend	Indep	Depend
1	1						51		(1)				
2		1					52		(1)				
3		2					53		(1)				
4		(1)					54		(1)				
5		(1)					55		(1)				
6		(1)					56		(1)				
7		(1)					57		(1)				
8		(1)					58		(1)				
9		(1)					59		(1)				
10		(1)					60		(1)				
11		(1)					61		(1)				
12		(1)					62		(1)				
13		(1)					63		(1)				
14		(1)					64		(1)				
15		(1)					65		(1)				
16		(1)					66		(1)				
17		(1)					67		(1)				
18		(1)					68		(1)				
19		(1)					69		(1)				
20		(1)					70		(1)				
21		(1)					71		(1)				
22		(1)					72		(1)				
23		(1)					73		(1)				
24		(1)					74		(1)				
25		(1)					75		(1)				
26		(1)					76		(1)				
27		(1)					77		(1)				
28		(1)					78		(1)				
29		(1)					79		(1)				
30		(1)					80		(1)				
31		(1)					81		(1)				
32		(1)					82		(1)				
33		(1)					83		(1)				
34		(1)					84		(1)				
35		(1)					85		(1)				
36		(1)					86		(1)				
37		(1)					87		(1)				
38		(1)					88	1					
39		(1)					89		1				
40		(1)					90		1				
41		(1)					91	1					
42		(1)					92	1					
43		(1)					93		1				
44		(1)					94		1				
45		(1)					95	1					
46		(1)					96		1				
47		(1)					97		2				
48		(1)					98		2				
49		(1)					99		(1)				
50		(1)					100		(1)				

MULTIPLE DEPENDENT CLAIM FEE CALCULATION SHEET

Substitute for Form PTO-1360
(For use with Form PTO/SB/06)

Application Number	Filing Date
15329683	

Applicant(s) Xiaowei Zhuang

* May be used for additional claims or amendments

CLAIMS	AS FILED		AFTER FIRST AMENDMENT		AFTER SECOND AMENDMENT	
	Indep	Depend	Indep	Depend	Indep	Depend
101		(1)				
102		(1)				
103		(1)				
104		(1)				
105		(1)				
106		(1)				
107		(1)				
108		(1)				
109		(1)				
110		(1)				
111		(1)				
112		(1)				
113		(1)				
114		(1)				
115		(1)				
116		(1)				
117		(1)				
118		(1)				
119		(1)				
120		(1)				
121		(1)				
122	1					
123		1				
124		2				
125		(1)				
126		(1)				
127		(1)				
128		(1)				
129		(1)				
130		(1)				
131		(1)				
132		(1)				
133		(1)				
134		(1)				
135		(1)				
136	1					
137		1				
138	1					
139		1				
140		2				
141	1					
142		(1)				
143		(1)				
144	1					
145		1				
146		2				
147		2				
148		(1)				
149		(1)				
150		(1)				
Total Indep	11		0		0	
Total Depend	176	↵	0	↵	0	↵
Total Claims	187		0		0	

	*		*		*	
	Indep	Depend	Indep	Depend	Indep	Depend
151		(1)				
152		(1)				
153		(1)				
154		(1)				
155		(1)				
156		(1)				
157		(1)				
158		(1)				
159		(1)				
160		(1)				
161		(1)				
162		(1)				
163		(1)				
164		(1)				
165		(1)				
166	1					
167		1				
168		1				
169		1				
170		1				
171		5				
172		(1)				
173		(1)				
174		(1)				
175		(1)				
176		(1)				
177						
178						
179						
180						
181						
182						
183						
184						
185						
186						
187						
188						
189						
190						
191						
192						
193						
194						
195						
196						
197						
198						
199						
200						

Docket No.: H0498.70509US02
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	Not Yet Assigned
Art Unit:	Not Yet Assigned

Certificate of Electronic Filing under 37 CFR §1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: June 13, 2017 Electronic Signature for Sarah Carletti: /Sarah Carletti/

PRELIMINARY AMENDMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

INTRODUCTORY COMMENTS

Prior to examination on the merits, please amend the above-identified U.S. patent application as indicated below.

The Claims are reflected in the listing of claims which begins on page 2 of this paper.

Remarks begin on page 7 of this paper.

IN THE CLAIMS

Please replace all prior versions, and listings, of claims in the application with the following list of claims. Additions are indicated by underlining and deletions are indicated by strikeouts and/or double bracketing.

1. (Original) A method, comprising:

exposing a sample to a plurality of nucleic acid probes;

for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample;

creating codewords based on the binding of the nucleic acid probes; and

for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

2-6. (Cancelled)

7. (Currently amended) The method of ~~claims 1-6~~ claim 1, wherein the plurality of nucleic acid probes comprises a combinatorial combination of nucleic acid probes with different sequences.

8-9. (Cancelled)

10. (Currently amended) The method of ~~any one of claims 1-9~~ claim 1, wherein at least some of the plurality of nucleic acid probes comprises a first portion comprising a target sequence and a second portion comprising one or more read sequences.

11-17. (Cancelled)

18. (Currently amended) The method of ~~any one of claims 10-17~~ claim 10, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

19-25. (Cancelled)

26. (Currently amended) The method of ~~claims 1-25~~ claim 1, further comprising exposing the sample to a first secondary probe comprising a first signaling entity, the first secondary probe able to bind to some of the read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the first signaling entity within the sample.

27-30. (Cancelled)

31. (Currently amended) The method of ~~any one of claims 26-30~~ claim 26, further comprising exposing the sample to a second secondary probe comprising a second signaling entity, the second secondary probe able to bind to some of read sequences of the nucleic acid probes, and determining binding of the nucleic acid probes by determining the second signaling entity within the sample.

32-33. (Cancelled)

34. (Currently amended) The method of ~~any one of claims 31-33~~ claim 31, further comprising inactivating the first signaling entity prior to exposing the sample to the second secondary probe.

35-43. (Cancelled)

44. (Currently amended) The method of ~~any one of claims 1-43~~ claim 1, comprising determining the centroid of the first signaling entity using an algorithm for determining non-overlapping single emitters.

45-49. (Cancelled)

50. (Currently amended) The method of ~~any one of claims 1-49~~ claim 1, further comprising determining a confidence level for the identified nucleic acid target.

51-70. (Cancelled)

71. (Currently amended) The method of ~~any one of claims 1-70~~ claim 1, comprising determining binding of the nucleic acid probes within the sample at a resolution better than 300 nm.

72-74. (Cancelled)

75. (Currently amended) The method of ~~any one of claims 1-74~~ claim 1, wherein the sample comprises a cell.

76. (Cancelled)

77. (Currently amended) The method of ~~any one of claims 76 or 77~~ claim 75, wherein the cell is fixed.

78. (Currently amended) The method of any ~~one of claims 1-77~~ claim 1, comprising determining binding of the nucleic acid probes by imaging at least a portion of the sample.

79-94. (Cancelled)

95. (Original) A composition, comprising:

a plurality of nucleic acid probes, each comprising a first portion comprising a target sequence and a plurality of read sequences, wherein the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

96. (Original) The composition of claim 95, wherein the plurality of nucleic acid probes defines a code space with a Hamming distance of at least 2.

97-98. (Cancelled)

99. (Currently amended) The composition of ~~any one of claims 95-98~~ claim 95, wherein the code is a SECDED code.

100-102. (Cancelled)

103. (Currently amended) The composition of ~~any one of claims 99-102~~ claim 99, wherein only codewords with a constant number of ones are used.

104-121. (Cancelled)

122. (Original) A method, comprising:

associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space;

associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and

forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

123-125. (Cancelled)

126. (Currently amended) A method, comprising:

exposing nucleic acid probes formed using the method of ~~any one of claims 122-125~~ claim 122 to a sample;

sequentially determining binding of each of a plurality of secondary nucleic probes to the nucleic acid probes within the sample;

determining sample codewords comprising a pattern of bindings of the plurality of secondary nucleic probes to the nucleic acid probes; and

matching the sample codewords to the codewords to determining targets within the sample.

127-128. (Cancelled)

129. (Currently amended) The method of claim ~~125-128~~ 126, wherein if no match is found between a valid codeword and the sample codeword, applying error correction to the sample codeword to identified a matched codeword.

130-176. (Cancelled)

REMARKS

This is a Preliminary Amendment in which claims 2-6, 8, 9, 11-17, 19-25, 27-30, 32, 33, 35-43, 45-49, 51-70, 72-74, 76, 79-94, 97, 98, 100-102, 104-121, 123-125, 127, 128, and 130-176 have been cancelled without prejudice or disclaimer. In addition, many of the remaining claims have been amended to eliminate multiple dependencies.

Please charge our Credit Card in the amount of $70.00. The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed, or which should have been filed herewith to Deposit Account No. 23/2825 under Docket No. H0498.70509US02, from which the undersigned is authorized to draw.

Dated: June 13, 2017

Respectfully submitted,

Electronic signature: /Tani Chen/
Tani Chen, Sc.D.
Registration No.: 52,728
Timothy J. Oyer, Ph.D.
Registration No.: 36,628
WOLF, GREENFIELD & SACKS, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
617.646.8000

POST AIA

DECLARATION (37 C.F.R. § 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 C.F.R. § 1.76)

Title of Invention	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

As the below named inventor, I declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number PCT/US2015/042556 filed on July 29, 2015 (amended January 27, 2017 – U.S. 15/329,683)

The above-identified application was made or authorized to be made by me.

I believe I am the original inventor or an original joint inventor of a claimed invention in the application.

I have reviewed and understand the contents of the above-identified application, including the claims.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 C.F.R. § 1.56.

All statements made herein of my own knowledge are true, and all statements made herein on information and belief are believed to be true.

I hereby acknowledge that any willful false statement and the like made in this declaration is punishable by fine or imprisonment of not more than five (5) years, or both, under 18 U.S.C. § 1001, and may jeopardize the validity of the application or any patent issuing thereon.

LEGAL NAME OF SOLE OR FIRST INVENTOR

Inventor one: Xiaowei Zhuang Date: 4/25/17

Signature: [signature]

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor two: Kok-Hao Chen Date:

Signature:

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor three: Alistair Boettiger Date:

Signature:

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor four: Jeffrey R. Moffitt Date: 4/25/17

Signature: [signature]

☒ Additional inventors are being named on 1 Additional form(s) attached hereto.

POST AIA

DECLARATION FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET	ADDITIONAL INVENTOR(S) Supplemental Sheet Page 1 of 1
LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any	
Inventor five: Siyuan Wang Date: 4.24.2017 Signature: Siyuan Liang	

DECLARATION (37 C.F.R. § 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 C.F.R. § 1.76)

Title of Invention	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

As the below named inventor, I declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number PCT/US2015/042556 filed on July 29, 2015 (amended January 27, 2017 – U.S. 15/329,683)

The above-identified application was made or authorized to be made by me.

I believe I am the original inventor or an original joint inventor of a claimed invention in the application.

I have reviewed and understand the contents of the above-identified application, including the claims.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 C.F.R. § 1.56.

All statements made herein of my own knowledge are true, and all statements made herein on information and belief are believed to be true.

I hereby acknowledge that any willful false statement and the like made in this declaration is punishable by fine or imprisonment of not more than five (5) years, or both, under 18 U.S.C. § 1001, and may jeopardize the validity of the application or any patent issuing thereon.

LEGAL NAME OF SOLE OR FIRST INVENTOR

Inventor one: Xiaowei Zhuang Date: _____

Signature: _____

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor two: Kok-Hao Chen Date: 25/04/2017

Signature: Chen

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor three: Alistair Boettiger Date: _____

Signature: _____

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor four: Jeffrey R. Moffitt Date: _____

Signature: _____

☒ Additional inventors are being named on 1 Additional form(s) attached hereto.

POST AIA

DECLARATION FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET	ADDITIONAL INVENTOR(S) Supplemental Sheet Page 1 of 1
LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any	
Inventor five: Siyuan Wang	Date:
Signature:	

POST AIA

DECLARATION (37 C.F.R. § 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 C.F.R. § 1.76)

Title of Invention	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

As the below named inventor, I declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number PCT/US2015/042556 filed on July 29, 2015 (amended January 27, 2017 – U.S. 15/329,683)

The above-identified application was made or authorized to be made by me.

I believe I am the original inventor or an original joint inventor of a claimed invention in the application.

I have reviewed and understand the contents of the above-identified application, including the claims.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 C.F.R. § 1.56.

All statements made herein of my own knowledge are true, and all statements made herein on information and belief are believed to be true.

I hereby acknowledge that any willful false statement and the like made in this declaration is punishable by fine or imprisonment of not more than five (5) years, or both, under 18 U.S.C. § 1001, and may jeopardize the validity of the application or any patent issuing thereon.

LEGAL NAME OF SOLE OR FIRST INVENTOR

Inventor one: Xiaowei Zhuang Date: _____

Signature: _____

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor two: Kok-Hao Chen Date: _____

Signature: _____

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor three: Alistair Boettiger Date: 5/3/17

Signature: Alistair Boettiger

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor four: Jeffrey R. Moffitt Date: _____

Signature: _____

☒ Additional inventors are being named on 1 Additional form(s) attached hereto.

DECLARATION FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET	ADDITIONAL INVENTOR(S) Supplemental Sheet Page 1 of 1

LEGAL NAME OF ADDITIONAL JOINT INVENTOR, if any

Inventor five: Siyuan Wang Date: _____

Signature: _____

FEE TRANSMITTAL

Complete if known	
Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Examiner Name	Not Yet Assigned
Art Unit	Not Yet Assigned
Practitioner Docket No.	H0498.70509US02

[x] Applicant asserts small entity status. See 37 CFR 1.27.

[ ] Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($)70.00

METHOD OF PAYMENT (check all that apply)

[ ] Check [x] Credit Card [ ] Money Order [ ] None [ ] Other (please identify): _____

[x] Deposit Account Deposit Account Number: 23/2825 Deposit Account Name: Wolf, Greenfield & Sacks, P.C.

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

[ ] Charge fee(s) indicated below [ ] Charge fee(s) indicated below, **except for the filing fee**

[x] Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 [x] Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	280	140*	70	600	300	150	720	360	180	
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims **Extra Claims** **Fee ($)** **Fee Paid ($)**

_____ - 20 or HP = _____ x _____ = _____

HP = highest number of total claims paid for, if greater than 20.

Multiple Dependent Claims **Fee ($)** _____ **Fee Paid ($)** _____

Indep. Claims **Extra Claims** **Fee ($)** **Fee Paid ($)**

_____ - 3 or HP = _____ x _____ = _____

HP = highest number of independent claims paid for, if greater than 3.

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets **Extra Sheets** **Number of each additional 50 or fraction thereof** **Fee ($)** **Fee Paid ($)**

_____ - 100 = _____ /50 = _____ (round **up** to a whole number) x _____ = _____

4. OTHER FEE(S) **Fees Paid ($)**

Non-English specification, $130 fee (no small or micro entity discount) _____

Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity) _____

Other (e.g., late filing surcharge): 2051 Surcharge - Late filing fee, search fee, ... 70.00

SUBMITTED BY

Signature	/Tani Chen/	Registration No. (Attorney/Agent)	52,728	Telephone	617.646.8000
Name (Print/Type)	Tani Chen, Sc.D.			Date	June 13, 2017

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: June 13, 2017 Electronic Signature for Sarah Carletti: /Sarah Carletti/

Electronic Patent Application Fee Transmittal

Application Number:	15329683
Filing Date:	27-Jan-2017
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Filer:	Tani Chen/sarah carletti
Attorney Docket Number:	H0498.70509US02

Filed as Small Entity

Filing Fees for U.S. National Stage under 35 USC 371

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
OATH/DECL > 30 MOS FROM 371 COMMENCEMENT	2617	1	70	70
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**70**

Electronic Acknowledgement Receipt

EFS ID:	29482376
Application Number:	15329683
International Application Number:	
Confirmation Number:	7428
Title of Invention:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Xiaowei Zhuang
Customer Number:	86110
Filer:	Tani Chen/sarah carletti
Filer Authorized By:	Tani Chen
Attorney Docket Number:	H0498.70509US02
Receipt Date:	13-JUN-2017
Filing Date:	27-JAN-2017
Time Stamp:	15:33:22
Application Type:	U.S. National Stage under 35 USC 371

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$70
RAM confirmation Number	061417INTEFSW15341000
Deposit Account	232825
Authorized User	Wolf Greenfield
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows: 37 CFR 1.21 (Miscellaneous fees and charges)	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	H049870509US02-TRN-TC.PDF	26104 72edb544519c5c08bf6a74db7cfd6f1872300520	no	1
Warnings:					
Information:					
2	Preliminary Amendment	H049870509US02-PAMN-TC.PDF	37371 76c515eb7b7280cc3cb9c4536203f3e494f419ac	no	7
Warnings:					
Information:					
3	Oath or Declaration filed	H049870509US02-INVDEC-TC.PDF	193031 725d737b5a8a7b58243eb28a65c2ab56a90b965d	no	6
Warnings:					
Information:					
4	Fee Worksheet (SB06)	H049870509US02-FEE-TC.PDF	32152 89781c1529058fb76b05c00a12a9f0d32ce48b96	no	1
Warnings:					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	30593 8da6a0f3163e9c3e05995b77bf34105792f42581	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	319251		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	

Application Number	15/329,683-Conf. #7428
Filing Date	January 27, 2017
First Named Inventor	Xiaowei Zhuang
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	H0498.70509US02

ENCLOSURES (*Check all that apply*)

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

  Preliminary Amendment
  Declaration

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	WOLF, GREENFIELD & SACKS, P.C.		
Signature	/Tani Chen/		
Printed name	Tani Chen, Sc.D.		
Date	June 13, 2017	Reg. No.	52,728

Certificate of Electronic Filing under 37 CFR §1.8

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 CFR § 1.6(a)(4).

Dated: June 13, 2017 Electronic Signature for Sarah Carletti: /Sarah Carletti/

5242801.1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02

CONFIRMATION NO. 7428

86110
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

PUBLICATION NOTICE

OC000000093215980

Date Mailed: 08/03/2017

Title:SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

Publication No.US-2017-0220733-A1
Publication Date:08/03/2017

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Public Records Division. The Public Records Division can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Public Records Division, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently https://portal.uspto.gov/pair/PublicPair. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/329,683	01/27/2017	1639	1030	H0498.70509US02	20	3

CONFIRMATION NO. 7428

UPDATED FILING RECEIPT

86110
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

OC000000097056021

Date Mailed: 02/01/2018

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Xiaowei Zhuang, Lexington, MA;
Kok-Hao Chen, Singapore, SINGAPORE;
Alistair Boettiger, Cambridge, MA;
Jeffrey R. Moffitt, Somerville, MA;
Siyuan Wang, Cambridge, MA;

Applicant(s)

President and Fellows of Harvard College, Cambridge, MA;

Assignment For Published Patent Application

President and Fellows of Harvard College, Cambridge, MA

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a 371 of PCT/US15/42556 07/29/2015
which claims benefit of 62/031,062 07/30/2014
and claims benefit of 62/050,636 09/15/2014
and claims benefit of 62/142,653 04/03/2015

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: No

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 04/20/2017
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/329,683**
Projected Publication Date: Not Applicable
Non-Publication Request: No
Early Publication Request: No
**** SMALL ENTITY ****
Title

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

Preliminary Class

506

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative,

this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 15/329,683

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	150		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	260		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	380		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	20 minus 20 =	*	x 50 =	0.00	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	3 minus 3 =	*	x 230 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			400			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	1190		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)	OR	RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

U.S. APPLICATION NUMBER NO.	FIRST NAMED INVENTOR	ATTY. DOCKET NO.
15/329,683	Xiaowei Zhuang	H0498.70509US02

86110
Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

INTERNATIONAL APPLICATION NO.	
PCT/US15/42556	
I.A. FILING DATE	PRIORITY DATE
07/29/2015	07/30/2014

CONFIRMATION NO. 7428
371 ACCEPTANCE LETTER

OC000000097056022

Date Mailed: 02/01/2018

NOTICE OF ACCEPTANCE OF APPLICATION UNDER 35 U.S.C 371 AND 37 CFR 1.495

The applicant is hereby advised that the United States Patent and Trademark Office, in its capacity as a Designated / Elected Office (37 CFR 1.495), has ACCEPTED the above identified international application for national patentability examination in the United States Patent and Trademark Office.

The United States Application Number assigned to the application is shown above. A Filing Receipt will be issued for the present application in due course. **THE DATE APPEARING ON THE FILING RECEIPT AS THE "FILING DATE or 371(c) DATE" IS THE DATE ON WHICH THE LAST OF THE 35 U.S.C. 371 (c)(1) and (c)(2) REQUIREMENTS HAS BEEN RECEIVED IN THE OFFICE. THIS DATE IS SHOWN BELOW.** The filing date of the above identified application is the international filing date of the international application (Article 11(3) and 35 U.S.C. 363)

01/27/2017
DATE OF RECEIPT OF 35 U.S.C.
371(c)(1) and (c)(2) REQUIREMENTS

The following items have been received:

- Indication of Small Entity Status
- Copy of the International Application filed on 01/27/2017
- Copy of the International Search Report filed on 01/27/2017
- Preliminary Amendments filed on 01/27/2017
- Information Disclosure Statements filed on 04/12/2017
- Sequence Listing txt file (EFS or physical medium) filed on 01/27/2017
- Inventor's Oath or Declaration filed on 06/13/2017
- U.S. Basic National Fees filed on 01/27/2017
- Priority Documents filed on 01/27/2017
- Specification filed on 01/27/2017
- Claims filed on 01/27/2017
- Abstracts filed on 01/27/2017
- Drawings filed on 01/27/2017
- Application Data Sheet (37 CFR 1.76) filed on 01/27/2017

Applicant is reminded that any communications to the United States Patent and Trademark Office must be mailed to the address given in the heading and include the U.S. application no. shown above (37 CFR 1.5)

BRIAN A HARDEN

Telephone: (703) 756-1815

FORM PCT/DO/EO/903 (371 Acceptance Notice)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/329,683	01/27/2017	Xiaowei Zhuang	H0498.70509US02	7428

86110 7590 09/24/2018

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206

EXAMINER
LU, FRANK WEI MIN

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
09/24/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

H0498_eOfficeAction@WolfGreenfield.com
Patents_eOfficeAction@WolfGreenfield.com

Office Action Summary	**Application No.** 15/329,683	**Applicant(s)** Zhuang et al.	
	Examiner FRANK W LU	**Art Unit** 1634	**AIA Status** Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 2 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 6/13/2018.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) See Continuation Sheet is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☐ Claim(s) _____ is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☑ Claim(s) See Continuation Sheet are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

Continuation of Disposition of Claims* 5) Claim(s) is/are pending in the application: 1,7,10,18,26,31,34,44,50,71,75,77-78,95-96,99,103,122,126 and 129

Continuation of Disposition of Claims* 9) Claim(s) are subject to restriction and/or election requirement: 1,7,10,18,26,31,34,44,50,71,75,77-78,95-96,99,103,122,126 and 129

DETAILED ACTION

Preliminary Amendment

1. The preliminary amendment filed on June 13, 2017 has been entered. The claims pending in this application are claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, 78, 95, 96, 99, 103, 122, 126, and 129.

Election/Restrictions

2. REQUIREMENT FOR UNITY OF INVENTION

As provided in 37 CFR 1.475(a), a national stage application shall relate to one invention only or to a group of inventions so linked as to form a single general inventive concept ("requirement of unity of invention"). Where a group of inventions is claimed in a national stage application, the requirement of unity of invention shall be fulfilled only when there is a technical relationship among those inventions involving one or more of the same or corresponding special technical features. The expression "special technical features" shall mean those technical features that define a contribution which each of the claimed inventions, considered as a whole, makes over the prior art.

The determination whether a group of inventions is so linked as to form a single general inventive concept shall be made without regard to whether the inventions are claimed in separate claims or as alternatives within a single claim. See 37 CFR 1.475(e).

When Claims Are Directed to Multiple Categories of Inventions:

As provided in 37 CFR 1.475 (b), a national stage application containing claims to different categories of invention will be considered to have unity of invention if the claims are drawn only to one of the following combinations of categories:

(1) A product and a process specially adapted for the manufacture of said product; or

(2) A product and a process of use of said product; or

(3) A product, a process specially adapted for the manufacture of the said product, and a use of the said product; or

(4) A process and an apparatus or means specifically designed for carrying out the said process; or

(5) A product, a process specially adapted for the manufacture of the said product, and an apparatus or means specifically designed for carrying out the said process.

Otherwise, unity of invention might not be present. See 37 CFR 1.475 (c).

3. Restriction is required under 35 U.S.C. 121 and 372.

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In accordance with 37 CFR 1.499, applicant is required, in reply to this action, to elect a single invention to which the claims must be restricted.

Group I, claims 1, 7, 10, 18, 26, 31, 34, 44, 50, 71, 75, 77, and 78, drawn to a method.

Group II, claims 95, 96, and 99, drawn to a composition.

Group III, claims 122, 126, and 129, drawn to a method.

4. The groups of inventions listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Groups I and III and Group II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features. For example, the special technical feature linking Groups I and III and Group

II is not special because the composition in claim 95 of Group II is not a contribution over the prior art (eg., see Figures 1 and 4 of US 2012/0208724 A1, published on August 16, 2012).

Groups I and III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features. For example, creating step in claim 1 of Group I is not required for Group III while determining sample codewords in claim 126 of Group III is not required for Group I.

Applicant is reminded that upon the cancellation of claims to a non-elected invention, the inventorship must be corrected in compliance with 37 CFR 1.48(a) if one or more of the currently named inventors is no longer an inventor of at least one claim remaining in the application. A request to correct inventorship under 37 CFR 1.48(a) must be accompanied by an application data sheet in accordance with 37 CFR 1.76 that identifies each inventor by his or her legal name and by the processing fee required under 37 CFR 1.17(i).

5. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent

Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
September 18, 2018

Notice of References Cited	Application/Control No. 15/329,683	Applicant(s)/Patent Under Reexamination Zhuang et al.	
	Examiner FRANK W LU	Art Unit 1634	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-20120208724-A1	08-2012	Steemers; Frank J.	C12Q1/6869	506/26
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 7428

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
15/329,683	01/27/2017	435	1634	H0498.70509US02

APPLICANTS

President and Fellows of Harvard College, Cambridge, MA;

INVENTORS

Xiaowei Zhuang, Lexington, MA;
Kok-Hao Chen, Singapore, SINGAPORE;
Alistair Boettiger, Cambridge, MA;
Jeffrey R. Moffitt, Somerville, MA;
Siyuan Wang, Cambridge, MA;

**** CONTINUING DATA ***************************

This application is a 371 of PCT/US15/42556 07/29/2015
which claims benefit of 62/031,062 07/30/2014
and claims benefit of 62/050,636 09/15/2014
and claims benefit of 62/142,653 04/03/2015

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED **** ** SMALL ENTITY **
04/20/2017

Foreign Priority claimed	❑ Yes ☑ No	❑ Met after Allowance	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
35 USC 119(a-d) conditions met	❑ Yes ☑ No		MA	48	20	3
Verified and Acknowledged	/FRANK WEI MIN LU/ Examiner's Signature	Initials				

ADDRESS

Wolf, Greenfield & Sacks, P.C. (Harvard)
600 Atlantic Avenue
Boston, MA 02210-2206
UNITED STATES

TITLE

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

FILING FEE RECEIVED		
1030	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	❑ All Fees ❑ 1.16 Fees (Filing) ❑ 1.17 Fees (Processing Ext. of time) ❑ 1.18 Fees (Issue) ❑ Other _____ ❑ Credit

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	1	of	2	GROUP ART UNIT: 1634	EXAMINER: Frank Wei Min Lu

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document		Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		Number	Kind Code		
		6,524,798	B1	Goldbard et al.	02-25-2003
		7,368,265	B2	Brenner et al.	05-06-2008
		7,838,302		Zhuang et al.	11-23-2010
		8,564,792	B2	Zhuang et al.	10-22-2013
		2004-0137488	A1	Kenten et al.	07-15-2004
		2006-0051798	A1	Mirken et al.	03-09-2006
		2008-0269068	A1	Church et al.	10-30-2008
		2015-0087001	A1	Gradinaru et al.	03-26-2015
		2016-0116384	A1	Chen et al.	04-28-2016
		2016-0304952	A1	Boyden et al.	10-20-2016
		2017-0212986	A1	Zhuang et al.	07-27-2018

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document			Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		Office/ Country	Number	Kind Code			
		WO	1999/031277	A1	Medical Science Systems, Inc.	06-24-1999	
		WO	03/003810	A3	Inst. Systems Biology	01-16-2003	
		WO	03/003810	A2	Inst. Systems Biology	01-16-2003	
		WO	2008/108843	A2	Gen Probe, Inc.	09-12-2008	
		WO	2013/090360	A2	President and Fellows of Harvard College	06-20-2013	
		WO	2016/018960	A1	President and Fellows of Harvard College	02-04-2016	

OTHER ART – NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		Extended European Search Report mailed February 6, 2018 for Application No. 15827358.1 (H0498.70509EP00).	
		European Office Action mailed September 26, 2018 for Application No. 15827358.1 (H0498.70509EP00).	
		Extended European Search Report mailed February 5, 2018 for Application No. 15828133.7 (H0498.70510EP00).	
		European Office Action mailed September 24, 2018 for Application No. 15828133.7 (H0498.70510EP00).	
		International Search Report and Written Opinion mailed February 5, 2018 for Application No. PCT/US17/60570 (H0498.70630WO00).	
		Invitation to Pay Additional Fees mailed August 6, 2018 for Application No. PCT/US18/34651 (H0498.70637WO00).	

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant

FORM PTO-1449/A and B (modified PTO/SB/08)				APPLICATION NO.: 15/329,683	ATTY. DOCKET NO.: H0498.70509US02
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				FILING DATE: January 27, 2017	CONFIRMATION NO.: 7428
				FIRST NAMED INVENTOR: Xiaowei Zhuang	
Sheet	2	of	2	GROUP ART UNIT: 1634	EXAMINER: Frank Wei Min Lu

		International Search Report and Written Opinion mailed October 4, 2018 for Application No. PCT/US18/34651 (H0498.70637WO00).	
		International Search Report and Written Opinion mailed February 5, 2018 for Application No. PCT/US17/60558 (H0498.70645WO00).	
		BINLADEN et al., The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One. 2007 Feb 14;2(2):e197.	
		CHEN et al., Expansion microscopy. Science. 2015 Jan; 347(6221):543-548	
		CHEN et al., Nanoscale imaging of RNA with expansion microscopy. Nature Methods. 2016; 13:679-684.	
		DAROUICH et al., Use of DOP-PCR for amplification and labeling of BAC DNA for FISH. Biotech Histochem. 2012 Feb;87(2):117-21. doi: 10.3109/10520295.2011.559175.	
		TILLBERG et al., Protein-retention expansion microscopy of cells and tissues labeled using standard flourescent proteins and antibodies. Nat. Biotechnol. 2016; 34(9):987-992.	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant

DOCKET NO.: H0498.70509US02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Xiaowei Zhuang
Application No.:	15/329,683
Confirmation No.:	7428
Filed:	January 27, 2017
For:	SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS
Examiner:	Frank Wei Min Lu
Art Unit:	1634

CERTIFICATE OF ELECTRONIC FILING UNDER 37 C.F.R. § 1.8

The undersigned hereby certifies that this paper, along with any paper referred to as being attached or enclosed, is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4), on the 20th day of November, 2018.

/Kimberly E. Martin/
Kimberly E. Martin

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action on the merits in the above-identified case.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of

presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
* 15/329,651	1/27/2017	Zhuang et al.	H0498.70510US01

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR.

The undersigned would like to bring to the Examiner's attention the enclosed search report or other communication from a corresponding or related International or Foreign National Application:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.
EP 15827358.1	**02-06-2018**	**Extended European Search Report**	**H0498.70509EP00**
EP 15827358.1	**09-26-2018**	**European Office Action**	**H0498.70509EP00**
EP 15828133.7	**02-05-2018**	**Extended European Search Report**	**H0498.70510EP00**
EP 15828133.7	**09-24-2018**	**European Office Action**	**H0498.70510EP00**
PCT/US17/60570	**02-05-2018**	**International Search Report and Written Opinion**	**H0498.70630WO00**
PCT/US18/34651	**08-06-2018**	**Invitation to Pay Additional Fees**	**H0498.70637WO00**
PCT/US18/34651	**10-04-2018**	**International Search Report and Written Opinion**	**H0498.70637WO00**
PCT/US17/60558	**02-05-2018**	**International Search Report and Written Opinion**	**H0498.70645WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. H0498.70509US02 from which the undersigned is authorized to draw.

Respectfully submitted,

By: /Tani Chen/
Tani Chen, Sc.D., Reg. No. 52,728
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: H0498.70509US02
Date: November 20, 2018
xNDDx

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference H0498-70630	Date of mailing *(day/month/year)* 05 FEB 2018 **FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 17/60570	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee W. Young PCT Helpdesk: 571-272-4300 Telephone No. PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2014)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference H0498-70630	FOR FURTHER ACTION	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 17/60570	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)	(Earliest) Priority Date *(day/month/year)* 08 November 2016 (08.11.2016)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 9 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 1B

☒ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

Form PCT/ISA/210 (first sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60570

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60570

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-64, 68-118, 122-150, 154-169
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60570

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 19/12 (2018.01)
CPC - G06F 19/12

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015/0087001 A1 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 26 March 2015 (26.03.2015) Claim 1; para [0016-0018]; para [0025]; para [0037]; para [0054]; para [0063]	1-4, 65-67
X --- Y	US 2016/0304952 A1 (MASSACHUSSETTS INSTITUTE OF TECHNOLOGY et al.,) 20 October 2016 (20.10.2016) Claim 1; para [0010-0011]; para [0029]; para [0037]; para [0050]	151-153 --- 119-121
Y	MOFFITT et al., High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci U S A. 27 Sep. 2016, Vol. 113, No. 39, pp 11046-51. Abstract	119-121

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 January 2018	05FEB 2018
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	05FEB 2018

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498-70630	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 17/60570	08 November 2017 (08.11.2017)	08 November 2016 (08.11.2016)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G06F 19/12 (2018.01)
CPC - G06F 19/12

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	10 January 2018	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 5-64, 68-118, 122-150, 154-169

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 5-64,68-118,122-150,154-169 are so unclear that no meaningful opinion could be formed *(specify)*:

Claims 5-64, 68-118, 122-150, 154-169 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 5-64, 68-118, 122-150, 154-169

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

Form PCT/ISA/237 (Box No. III) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

I. Statement

Novelty (N)	Claims	4, 66, 67, 119-121	YES
	Claims	1-3, 65, 151-153	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-4, 65-67, 119-121, 151-153	NO
Industrial applicability (IA)	Claims	1-4, 65-67, 119-121, 151-153	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-3 and 65 lack novelty under PCT Article 33(2) as being anticipated by US 2015/0087001 A1 to California Institute Of Technology (hereinafter 'Caltech').

Regarding Claim 1, Caltech teaches a method, comprising:
exposing a sample to a plurality of nucleic acid probes (para [0017] - '100 um-thick mouse brain slices were hybridized with twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594.');
polymerizing a gel within the sample (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0025] - 'Immediately prior to hydrogel polymerization, the entire perfusion chamber (with catheter-connected subject) is sealed within a ziplock bag for tissue degassing');
anchoring a target to the gel; clearing non-targets from the sample (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0016] - 'FIG. 1 demonstrates, in accordance with an embodiment of the invention, PACT clearing of A4P0 tissue-hydrogel hybrid achieves optimal transparency and immunohistochemistry compatibility across organs.'; [0056] 'native proteins including those maintaining the structural integrity of tissue samples, retained by the tissue-hydrogel matrix during clearing'); and
determining the targets within the gel by determining binding of the nucleic acid probes by imaging (para [0017] - 'twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594...An abundant number of diffraction limited spots corresponding to single beta-actin mRNAs (red) were readily detected up to 30 um in depth under 589 nm illumination....the images were processed from raw data'; [0018] - ' After 24-hour incubation in RIMS, PARS-CSF brain slices were sufficiently cleared for imaging without further sectioning....Images show native eGFP fluorescence in 500 um PARS-CSF cleared coronal brain slice').

Regarding Claim 2, Caltech teaches the method of claim 1, wherein the target is a nuclelc acid (para [0017] - 'FIG. 2 demonstrates, in accordance with an embodiment of the invention, detection of individual mRNA transcripts in PACT tissue sections by smFISH.').

Regarding Claim 3, Caltech teaches the method of any one of claims 1 or 2, wherein the target comprises RNA (para [0017] - 'FIG. 2 demonstrates, in accordance with an embodiment of the invention, detection of individual mRNA transcripts in PACT tissue sections by smFISH.').

Regarding Claim 65, Caltech teaches a method, comprising:
exposing a sample to a plurality of nucleic acid probes (para [0017] - '100 um-thick mouse brain slices were hybridized with twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594.');
polymerizing a gel within the sample (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0025] - 'Immediately prior to hydrogel polymerization, the entire perfusion chamber (with catheter-connected subject) is sealed within a ziplock bag for tissue degassing');
anchoring a target to the gel (claim 1 - 'applying a hydrogel monomer solution comprising acrylamide and phosphate buffered saline (PBS) to the fixed tissue, thereby forming hydrogel treated tissue.'; para [0016] - 'FIG. 1 demonstrates, in accordance with an embodiment of the invention, PACT clearing of A4P0 tissue-hydrogel hybrid achieves optimal transparency and immunohistochemistry compatibility across organs.'; [0056] 'native proteins including those maintaining the structural integrity of tissue samples, retained by the tissue-hydrogel matrix during clearing');
reducing background fluorescence within the sample (para [0017] - '(B) Compared to PACT cleared slices, smFISH in uncleared brain slices showed significantly decreased contrast....Comparison of background intensity between uncleared and PACT-cleared tissue illustrates the significant reduction of background fluorescence in PACT-cleared tissue'); and
imaging the nucleic acid probes (para [0017] - 'twenty-four 20 mer oligonucleotide probes towards beta-actin mRNA labeled with Alexafluor 594...An abundant number of diffraction limited spots corresponding to single beta-actin mRNAs (red) were readily detected up to 30 um in depth under 589 nm illumination....the images were processed from raw data'; [0018] - ' After 24-hour incubation in RIMS, PARS-CSF brain slices were sufficiently cleared for imaging without further sectioning....Images show native eGFP fluorescence in 500 um PARS-CSF cleared coronal brain slice').

---Please see continuation in first supplemental box----------

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V2 Citations and Explanations

Claims 151-153 lack novelty under PCT Article 33(2) as being anticipated by US 2016/0304952 A1 to Massachusetts Institue Of Technology et al., (hereinafter 'MIT').

Regarding Claim 151, MIT teaches a method, comprising:
embedding at least a portion of a sample within a matrix (Claim 1 - 'b) embedding the biological sample comprising the target nucleic acids and attached small molecule linker or nucleic acid adaptor in a swellable material wherein the small molecule linker or the nucleic acid adaptor is linked to the target nucleic acids present in the sample and to the swellable material'; para [0037] - 'Monomers or oligomers can comprise one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides,');
immobilizing targets to the matrix (Claim 1 - 'a) linking target nucleic acids present in the biological sample with a small molecule linker or a nucleic acid adaptor capable of linking to a target nucleic acid and to a swellable material'; para [0029] - 'Additionally, acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel');
clearing non-targets from the matrix (Claim 1 - 'c) digesting proteins present in the biological sample'; para [0047] 'The degree of digestion can be sufficient to compromise the integrity of the mechanical structure of the sample or it can be complete to the extent that the sample-swellable material complex is rendered substantially free of the sample. Preferably the disruption of the physical structure of the sample is protein digestion of the proteins contained in the biological sample'); and imaging the targets within the matrix (para [0050] - 'The swollen material with the embedded biological sample can be imaged on any optical microscope'; para [0011] - 'FIG. 3 shows microscopy images showing in situ sequencing').

Regarding Claim 152, MIT teaches the method of claim 151, wherein the matrix comprises a polymer (para [0037] - 'Monomers or oligomers can comprise one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides,').

Regarding Claim 153, MIT teaches the method of any one of claims 151 or 152, wherein the matrix comprises a gel (para [0010] - 'Panel B depicts that biological specimens are embedded in a swellable gel material').

Claims 4, 66-67 lack an inventive step under PCT Article 33(3) as being obvious over Caltech.

Regarding Claim 4, Caltech teaches the method of any one of claims 1 or 2, yet does not expressly teach the target comprises DNA. Since Caltech teaches the method is intended for detection of DNA among other analytes (para [0037] - 'Thus motivation to improve tissue clearing protocols is sustained around three main objectives: 1) efficient clearing of both central organs and peripheral tissues; 2) preservation of cellular and subcellular structures of multiple organ types; and 3) compatibility with endogenous fluorescent protein expression and post-hoc detection of DNA, RNA, and proteins.'), it would have been obvious to one of ordinary skill in the art that the method of Caltech could be used to detect target DNA.

Regarding Claim 66, Caltech teaches the method of claim 65, yet does not expressly teach wherein during imaging, the gel has not expanded by more than 3x. Caltech does teach limited hydrogel swelling (para [0063] - 'Indeed, PARS-based clearing of rodent brains was accomplished with limited hydrogel swelling and tissue expansion during clearing (FIG. 4C-4D, 11A)'). Since Caltech teaches tissue-hydrogel matrices exhibited less than 300% (i.e., less than 3x) tissue weight and volume changes during imaging (para [0054] - 'Indeed, tissue-hydrogel matrices that were prepared using acrylamide alone (A4P0) exhibited tissue weight and volume changes of 174% and 223%, respectively (FIG. 1E) over A4P4 counterparts'), it would have been obvious to one of ordinary skill in the art that the gel had not expanded by more than 3x since the overall tissue-hydrogel matrices exhibited less than 3x expansion.

Regarding Claim 67, Caltech teaches the method of any one of claims 65 or 66, yet does not expressly teach wherein during imaging, the gel has not expanded by more than 1.5x. Caltech does teach limited hydrogel swelling (para [0063] - 'Indeed, PARS-based clearing of rodent brains was accomplished with limited hydrogel swelling and tissue expansion during clearing (FIG. 4C-4D, 11A)'). Since Caltech teaches tissue-hydrogel matrices exhibited a variable range of swelling based on their reagent composition (para [0054] - 'Indeed, tissue-hydrogel matrices that were prepared using acrylamide alone (A4P0) exhibited tissue weight and volume changes of 174% and 223%, respectively (FIG. 1E) over A4P4 counterparts'), it would have been obvious to one of ordinary skill in the art that by varying the reagent composition such as the percentage of paraformaldehyde vs acrylamide, one could identify by routine experimentation a reagent composition wherein the gel expansion does not exceed 1.5x.

----Please see continuation in next supplemental box----------

Form PCT/ISA/237 (Supplemental Box) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60570

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:
Box V2 Citations and Explanations

Claims 119-121 lack an inventive step under PCT Article 33(3) as being obvious over MIT in view of the article entitled 'High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization' by Moffitt et al., (hereinafter 'Moffitt').

Regarding Claim 119, MIT teaches a method (Claim 1 - 'A method for in-situ sequencing of target nucleic acids present in a biological sample'), comprising:
exposing a sample to a plurality of anchor nucleic acid probes (Claim 1 - 'a) linking target nucleic acids present in the biological sample with a small molecule linker or a nucleic acid adaptor capable of linking to a target nucleic acid and to a swellable material'; para [0029] - ' For example, anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases, such as T4 RNA ligase, or may be attached via a chemical linker such as a reactive amine group capable of reacting with target nucleic acid.');
embedding at least a portion of the sample within a polyacrylamide gel (Claim 1 - 'b) embedding the biological sample comprising the target nucleic acids and attached small molecule linker or nucleic acid adaptor in a swellable material wherein the small molecule linker or the nucleic acid adaptor is linked to the target nucleic acids present in the sample and to the swellable material'; para [0037] - 'Monomers or oligomers can comprise one or more substituted or unsubstituted methacrylates, acrylates, acrylamides, methacrylamides,')
immobilizing at least some of the anchor nucleic acid probes to the polyacrylamide gel (para [0029] - 'Additionally, acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel'); clearing proteins and/or lipids and/or DNA and/or extracellular matrix and/or RNA molecules from the sample (Claim 1 - 'c) digesting proteins present in the biological sample'; para [0047] 'The degree of digestion can be sufficient to compromise the integrity of the mechanical structure of the sample or it can be complete to the extent that the sample-swellable material complex is rendered substantially free of the sample. Preferably the disruption of the physical structure of the sample is protein digestion of the proteins contained in the biological sample'); and determining sequence of target mRNA by imaging the polyacrylamide gel (para [0050] - 'The swollen material with the embedded biological sample can be imaged on any optical microscope'; para [0011] - 'FIG. 3 shows microscopy images showing in situ sequencing'). MIT does not expressly teach exposing a sample to a plurality of MERFISH nucleic acid probes; or determining binding of the MERFISH nucleic acid probes. Moffitt teaches exposing a sample to a plurality of MERFISH nucleic acid probes and determining binding of the MERFISH nucleic acid probes (Abstract - '. Here we report advances that lead to a drastic increase in the measurement throughput of multiplexed error-robust fluorescence in situ hybridization (MERFISH), an image based approach to single-cell transcriptomics. In MERFISH, RNAs are identified via a combinatorial labeling approach that encodes RNA species with error-robust barcodes followed by sequential rounds of single-molecule fluorescence in situ hybridization (smFISH) to read out these barcodes.'). Since both MIT and Moffitt teach methods based on smFISH, it would have been obvious to one of ordinary skill in the art that the MERFISH probes of Moffitt could be used with the method of MIT to image target mRNAs in situ.

Regarding Claim 120, MIT in view of Moffitt teaches the method of claim 119, wherein MIT further teaches the polyacrylamide gel comprises anchor probes incorporated within the polyacrylamide gel (para [0029] - 'Additionally, acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel').

Regarding Claim 121, MIT in view of Moffitt teaches the method of any one of claims 119 or 120, wherein MIT further teaches clearing proteins and/or lipids from the sample comprises removing proteins and/or lipids from the sample (Claim 1 - 'c) digesting proteins present in the biological sample').

Claims 1-4, 65-67, 119-121 and 151-153 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

PATENT COOPERATION TREATY

AUG 10 2018

Docketed ✓ Already Docketed
Not Required ✓
Initials 1st JW 2nd

From the INTERNATIONAL SEARCHING AUTHORITY

PCT IRD

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

INVITATION TO PAY ADDITIONAL FEES AND, WHERE APPLICABLE, PROTEST FEE

(PCT Article 17(3)(a) and Rules 40.1 and 40.2(e))

Date of mailing *(day/month/year)*	06 AUG 2018
Applicant's or agent's file reference H0498.70637	**PAYMENT DUE** within **ONE MONTH** from the above date of mailing
International application No. PCT/US18/34651	International filing date *(day/month/year)* 25 May 2018 (25.05.2018)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

I. This International Searching Authority

(i) considers that there are 2 *(number of)* inventions claimed in the international application covered by the claims indicated below/on an extra sheet:

Group I, Claims 1-4, 142-144, 149-154, 159, 161-164, 180-181
Group II, Claim 160

(ii) therefore considers that **the international application does not comply with the requirement of unity of invention** (Rules 13.1, 13.2 and 13.3) for the reasons indicated below/on an extra sheet:

-***-Please See Supplemental Page-***-

(iii) ☐ has carried out a partial international search (see Annex) ☒ will establish the international search report on those parts of the international application which relate to the invention first mentioned in claims Nos.:

Group I, Claims 1-4, 142-144, 149-154, 159, 161-164, 180-181

(iv) will establish the international search report on the other parts of the international application only if, and to the extent to which, additional fees are paid.

2. Consequently, the applicant is hereby **invited to pay**, within the time limit indicated above, additional fees in the amount indicated below:

$2080.00 (Fee per additional invention) x 1 (number of additional inventions) = $2,080.00 (See Item 2 of annex) (total amount of additional fees/currency)

3. The applicant is informed that, according to Rule 40.2(c), **the payment of any additional fees may be made under protest**, that is, a reasoned statement to the effect that the international application complies with the requirement of unity of invention or that the amount of the required additional fees is excessive, where applicable, subject to the payment of a protest fee.

Where the applicant pays additional fees under protest, the applicant is hereby invited, within the time limit indicated above, to pay a protest fee (Rule 40.2(e)) in the amount of _____ *(amount/currency)*

Where the applicant has not, within the time limit indicated above, paid the required protest fee, the protest will be considered not to have been made and the International Searching Authority will so declare.

4. ☒ Claim(s) Nos. 5-141, 145-148, 155-158, 165-179, 182-201 have been found to be unsearchable under Article 17(2)(b) because of defects under Article 17(2)(a) and therefore have not been included with any invention.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300

Form PCT/ISA/206 (April 2005)

INVITATION TO PAY ADDITIONAL FEES
AND, WHERE APPLICABLE, PROTEST FEE

International application No.
PCT/US18/34651

Item 2 (continued). For International Applications filed on or after 01 January 2014, Applicant is reminded that the search fee per additional invention indicated in item 2 is the undiscounted fee per additional invention. An Applicant may pay the search fee per additional invention fee reduced by 50% (small entity assertion) or 75% (micro entity certification), as appropriate. See 37 CFR 1.27 and 1.29.

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-4, 142-144, 149-154, 159, 161-164, 180-181 are directed toward introducing, into a plurality of cells, nucleic acids comprising an identification portion and an expression portion and imaging cells, for example, to determine phenotypes and/or genotypes in populations of cells.

Group II, Claim 160 is directed toward a YFAST protein.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include nucleic acid tags and probes, not present in Group II; the special technical features of Group II include SEQ ID NO: 3, not present in Group I.

Groups I and II share the technical features including: a detection tag.

However, these shared technical features are previously disclosed by WO 2016/018960 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) (hereinafter 'Harvard').

Harvard discloses a detection tag (page 46, lines 4-10; figure 1B).

Since none of the special technical features of the Groups I and II inventions are found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Harvard reference, unity of invention is lacking.

RECEIVED
Wolf, Greenfield & Sacks, P.C.

PATENT COOPERATION TREATY

IF OCT 04 2018

Docketed ✓ Already Docketed
Not Required
Initials 1st 2nd

From the INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 01 OCT 2018
Applicant's or agent's file reference H0498.70637	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US18/34651	International filing date *(day/month/year)* 25 May 2018 (25.05.2018)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450	Authorized officer Shane Thomas
Facsimile No. 571-273-8300	Telephone No. PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2017)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference H0498.70637	Date of mailing *(day/month/year)* **FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US18/34651	International filing date *(day/month/year)* 25 May 2018 (25.05.2018)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

Filing of amendments and statement under Article 19:
The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   - ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   - ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2017)

RECEIVED
Wolf, Greenfield & Sacks, P.C.

OCT 04 2018

Docketed ✓ Already Docketed ____
Not Required ____
Initials 1st ____ 2nd ____

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference H0498.70637	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US18/34651	International filing date *(day/month/year)* 25 May 2018 (25.05.2018)	(Earliest) Priority Date *(day/month/year)* 26 May 2017 (26.05.2017)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 5 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

   c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☒ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

   a. the figure of the **drawings** to be published with the abstract is Figure No. 1C

   ☒ as suggested by the applicant.

   ☐ as selected by this Authority, because the applicant failed to suggest a figure.

   ☐ as selected by this Authority, because this figure better characterizes the invention.

   b. ☐ none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US18/34651

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US18/34651

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-141, 145-148, 155-158, 165-179, 182-201
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

-***-Please See Supplemental Page-***-

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, 180-181

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US18/34651

A. CLASSIFICATION OF SUBJECT MATTER
IPC - C40B 30/04; C12Q 1/68; G06F 19/20, 19/22, 19/24 (2018.01)
CPC - C12N 15/1065, 15/1079; C12Q 1/68, 1/6813, 1/6841; G06F 19/20, 19/22, 19/24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 1999/031277 A1 (MEDICAL SCIENCE SYSTEMS, INC.) June 24, 1999; abstract; page 10, lines 13-18; page 17, lines 4-6; page 23, lines 12-16	150 --- 1-3, 4/2-3, 142-143, 144/142-143, 151-153, 154/152-153, 159, 163, 164/162-163
X --- Y	US 2006/0051798 A1 (MIRKIN et al.) March 9, 2006; abstract; paragraphs [0004], [0087], [0112], [0115]; claim 1	161-162, 180-181 --- 149, 164/162-163,
Y	US 6,524,798 B1 (GOLDBARD et al.) February 25, 2003; abstract; column 11, lines 33-34; claim 1	1-3, 4/2-3
Y	WO 2016/018960 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) February 4, 2016; abstract; page 14, lines 22-26; page 17, lines 18-20; page 56, lines 8-10; page 65, lines 1-3	142-143, 144/142-143, 151-153, 154/152-153, 159
Y	US 2008/0269068 A1 (CHURCH et al.) October 30, 2008; paragraph [0063]	149
Y	US 2004/0137488 A1 (KENTEN et al.) July 15, 2004; paragraphs [0018], [0037]	3, 4/3

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 September 2018 (18.09.2018)	01 OCT 2018
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US18/34651

-***-Continued from Box No. III Observations where unity of invention is lacking: -***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-4, 142-144, 149-154, 159, 161-164, 180-181 are directed toward introducing, into a plurality of cells, nucleic acids comprising an identification portion and an expression portion and imaging cells, for example, to determine phenotypes and/or genotypes in populations of cells.

Group II, Claim 160 is directed toward a YFAST protein.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include nucleic acid tags and probes, not present in Group II; the special technical features of Group II include SEQ ID NO: 3, not present in Group I.

Groups I and II share the technical features including: a detection tag.

However, these shared technical features are previously disclosed by WO 2016/018960 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) (hereinafter 'Harvard').

Harvard discloses a detection tag (page 46, lines 4-10; figure 1B).

Since none of the special technical features of the Groups I and II inventions are found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Harvard reference, unity of invention is lacking.

RECEIVED
Wolf, Greenfield & Sacks, P.C.
OCT 04 2018
Docketed ✓ Already Docketed
Not Required
Initials 1st 2nd

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: Tani Chen
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
United States of America

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	01 OCT 2018
Applicant's or agent's file reference H0498.70637	**FOR FURTHER ACTION** See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US18/34651	25 May 2018 (25.05.2018)	26 May 2017 (26.05.2017)

International Patent Classification (IPC) or both national classification and IPC
IPC - C40B 30/04; C12Q 1/68; G06F 19/20, 19/22, 19/24 (2018.01)
CPC - C12N 15/1065, 15/1079; C12Q 1/68, 1/6813, 1/6841; G06F 19/20, 19/22, 19/24

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	18 September 2018 (18.09.2018)	Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No
PCT/US18/34651

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 5-141, 145-148, 155-158, 165-179, 182-201

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. -***-Please See Below-***- are so unclear that no meaningful opinion could be formed *(specify)*:

Claims Nos.: 5-141, 145-148, 155-158, 165-179, 182-201 are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 5-141, 145-148, 155-158, 165-179, 182-201

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees the applicant has, within the applicable time limit:

   ☐ paid additional fees.

   ☐ paid additional fees under protest and, where applicable, the protest fee.

   ☐ paid additional fees under protest but the applicable protest fee was not paid.

   ☒ not paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-4, 142-144, 149-154, 159, 161-164, 180-181 are directed toward introducing, into a plurality of cells, nucleic acids comprising an identification portion and an expression portion and imaging cells, for example, to determine phenotypes and/or genotypes in populations of cells.

Group II, Claim 160 is directed toward a YFAST protein.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include nucleic acid tags and probes, not present in Group II; the special technical features of Group II include SEQ ID NO: 3, not present in Group I.

Groups I and II share the technical features including: a detection tag.

However, these shared technical features are previously disclosed by WO 2016/018960 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) (hereinafter 'Harvard').

Harvard discloses a detection tag (page 46, lines 4-10; figure 1B).

Since none of the special technical features of the Groups I and II inventions are found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Harvard reference, unity of invention is lacking.

4. Consequently, this opinion has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. -***-Please See Above-***-

Form PCT/ISA/237 (Box No. IV) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US18/34651

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	-***-Please See Below-***-	YES
	Claims	150, 161-162, 180-181	NO
Inventive step (IS)	Claims	NONE	YES
	Claims	-***-Please See Below-***-	NO
Industrial applicability (IA)	Claims	-***-Please See Below-***-	YES
	Claims	NONE	NO

2. Citations and explanations:

-Continued from Box No. V: 1. Statement.

Novelty (N): YES: Claims 1-3, 4/2-3, 142-143, 144/142-143, 149, 151-153, 154/152-153, 159, 163, 164/162-163, 180-181

Inventive step (IS): NO: Claims 1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, 180-181

Industrial Applicability (IA): YES: Claims 1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, 180-181

-Continued from Box No. V. 2: Citations and Explanations:

Claim 150 lacks novelty under PCT Article 33(2) as being anticipated by WO 1999/031277 A1 (Medical Science Systems, Inc.) (hereinafter 'MSI').
As per claim 150, MSI discloses a method (abstract), comprising: acquiring an image of a plurality of cells transfected with a plurality of nucleic acids comprising an identification portion and an expression portion (performing confocal microscopy on a plurality of cells transfected with a test promoter coupled with a reporter; abstract; page 14, lines 15-25), wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion (wherein cells fluoresce (wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion); page 14, line 15 – page 15, line 7); and acquiring a plurality of images of the plurality of cells (page 14, lines 15-25; figure 2), wherein the images of the cells exhibit differences due to differences in the identification portions of the nucleic acids within the cells (wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion); page 14, line 15 – page 15, line 7).

Claims 161-162 and 180-181 lack novelty under PCT Article 33(2) as being anticipated by US 2006/0051798 A1 to Mirkin et al. (hereinafter 'Mirkin').

As per claim 161, Mirkin discloses a method (abstract; claim 1), comprising: creating plurality of nucleic acids comprising an identification portion and an expression portion (creating plurality of nucleic acids comprising a DNA barcode (an identification portion) and a specific binding complement to a target analyte (an expression portion); paragraph [0162]; claim 1), wherein when expressed in cells, the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes (wherein when the target analyte is within a cell (expressed in cells), the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes; paragraph [0087]; claim 1).

As per claim 162, Mirkin discloses the method of claim 161, and Mirkin further discloses pairing an identification portion and an expression portion to form a nucleic acid (the nucleic acid comprises a DNA barcode (an identification portion) having some nucleic acids complementary to a specific binding complement to a target analyte (an expression portion); claim 1).

As per claim 180, Mirkin discloses a composition (paragraph [0004]; claim 1), comprising: a plurality of nucleic acids comprising an identification portion and an expression portion (comprising a plurality of nucleic acids comprising a DNA barcode (an identification portion) and a specific binding complement to a target analyte (an expression portion); paragraph [0132]; claim 1), wherein when expressed in cells, the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes (wherein when the target analyte is within a cell (expressed in cells), the identification portion is identifiable by sequentially exposing the cells to nucleic acid probes; paragraph [0087]; claim 1).

As per claim 181, Mirkin discloses the composition of claim 180, and Mirkin further discloses wherein the identification portion of the plurality of nucleic acids comprises at least 10 unique sequences (wherein the identification portion of the barcode comprises 4^12 different sequences (at least 10 unique sequences); paragraph [0112]).

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Box No. V: Citations and Explanations: -***-

Claims 1-2, and 4/2 lack an inventive step under PCT Article 33(3) as being obvious over US 6,524,798 B1 to Goldbard et al. (hereinafter 'Goldbard') in view of MSI.

As per claim 1, Goldbard discloses a method (abstract; claim 1), comprising: introducing, into a plurality of cells, antibodies comprising an identification portion (introducing into a plurality of cells an antibody bound to a ligand, wherein the antibody is specific for an antigen suspected of being associated with the cell (an identification portion); abstract; column 8, lines 33-35; claim 1); determining phenotype of the cells (abstract; claim 1); and determining genotype of the plurality of cells by sequentially exposing the plurality of cells to nucleic acid probes (determining the genotype of the cells by sequentially contacting the cells with polynucleotide probes; claim 1), and determining binding of target sequences of the nucleic acid probes within the plurality of cells (detecting the presence of the complex of the polynucleotide probe and cellular nucleic acid, and correlating the presence of said complex of the polynucleotide probe and cellular nucleic acid with the presence of the target sequence in the cell, thereby determining the cell genotype; claim 1). Goldbard does not disclose introducing nucleic acids comprising an identification portion and an expression portion. MSI discloses a method for detecting a phenotype (abstract), comprising introducing, into a plurality of cells, nucleic acids comprising an identification portion and an expression portion (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include introducing nucleic acids comprising an identification portion and an expression portion, as previously disclosed by MSI, in order to provide a superior nucleic acid that can be identified based on its identification portion.

As per claim 2, Goldbard and MSI, in combination, disclose the method of claim 1, but Goldbard does not disclose wherein introducing the nucleic acid comprises transfecting the nucleic acid into the plurality of cells. MSI discloses wherein introducing the nucleic acid comprises transfecting the nucleic acid into the plurality of cells (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include wherein introducing the nucleic acid comprises transfecting the nucleic acid into the plurality of cells, as previously disclosed by MSI, in order to provide a superior nucleic acid that can be identified based on its identification portion.

As per claim 4/2, Goldbard and MSI, in combination, disclose the method of claim 2, and Goldbard further discloses wherein the transfected cells are sorted (labeled cells are detected by FACS analysis; column 11, lines 33-34). Goldbard does not disclose transfected cells. MSI discloses transfected cells (abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include transfected cells, as previously disclosed by MSI, in order to provide a superior cell that has been transfected for easy identification.

Claims 142-143, 144/142-143, 151-153, 154/152-153, and 159 lack an inventive step under PCT Article 33(3) as being obvious over MSI in view of WO 2016/018960 A1 (President And Fellows Of Harvard College) (hereinafter 'Harvard').

As per claim 142, MSI discloses a method (abstract), comprising: introducing, into a plurality of cells, a nucleic acid comprising an identification portion and an expression portion (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract); imaging the plurality of cells to determine phenotype of the cells (determining the phenotype by FACS; page 10, lines 13-18). MSI does not disclose and determining genotype of the plurality of cells. Harvard discloses a method (abstract), comprising introducing, into a plurality of cells, a nucleic acid (a plurality of nucleic acid probes may be applied to a cell sample, and their binding within the sample determined using fluorescence, to determine locations of the nucleic acid probes within the sample; abstract) and determining genotype of the plurality of cells (determining the entire genome of the cell; page 17, lines 18-20). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include genotype of the plurality of cells, as previously disclosed by Harvard, in order to provide an improved method for identifying a cell based on phenotype and genotype.

As per claim 143, MSI and Harvard, in combination, disclose the method of claim 142, but MSI does not disclose wherein determining genotype comprises using MERFISH. Harvard discloses wherein determining genotype comprises using MERFISH (page 17, lines 18-20; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using MERFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

As per claims 144/142-143, MSI and Harvard, in combination, disclose the method of claims 142-143, but MSI does not disclose wherein determining genotype comprises using smFISH. Harvard discloses wherein determining genotype comprises using smFISH (gene expression is determined by MERFISH, which is a highly multiplexed smFISH imaging method; page 17, lines 18-20; page 56, lines 8-10; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using smFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Previous Supplemental Box-***-

As per claim 151, MSI discloses a method (abstract) comprising: acquiring an image of a plurality of cells transfected with a plurality of nucleic acids comprising an identification portion and an expression portion (performing confocal microscopy on a plurality of cells transfected with a test promoter coupled with a reporter; abstract; page 14, lines 15-25), wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion (wherein cells fluoresce (wherein the cells exhibit imagable differences in phenotype due to expression of the expression portion); page 14, line 15 – page 15, line 7); and performing FACS (page 17, lines 4-6). MSI does not disclose and performing in situ hybridization, FISH, multiplexed FISH, smFISH, CASFISH, and/or MERFISH on the cells. Harvard discloses performing MERFISH (page 17, lines 18-20; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include performing MERFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the phenotype of a cell.

As per claim 152, MSI discloses a method (abstract), comprising: acquiring an image of a plurality of cells transfected with a plurality of nucleic acids comprising an identification portion and an expression portion (performing FACS on a plurality of cells transfected with a test promoter coupled with a reporter; abstract; page 10, lines 13-18), and determining genotype of the plurality of cells. MSI does not disclose and determining genotype of the plurality of cells. Harvard discloses a method (abstract), comprising introducing, into a plurality of cells, a nucleic acid (a plurality of nucleic acid probes may be applied to a cell sample, and their binding within the sample determined using fluorescence, to determine locations of the nucleic acid probes within the sample; abstract) and determining genotype of the plurality of cells (determining the entire genome of the cell; page 17, lines 18-20). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include genotype of the plurality of cells, as previously disclosed by Harvard, in order to provide an improved method for identifying a cell based on genotype.

As per claim 153, MSI and Harvard, in combination, disclose the method of claim 152, but MSI does not disclose wherein determining genotype comprises using MERFISH. Harvard discloses wherein determining genotype comprises using MERFISH (page 17, lines 18-20; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using MERFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

As per claims 154/152-153, MSI and Harvard, in combination, disclose the method of claims 152-153, but MSI does not disclose wherein determining genotype comprises using smFISH. Harvard discloses wherein determining genotype comprises using smFISH (gene expression is determined by MERFISH, which is a highly multiplexed smFISH imaging method; page 17, lines 18-20; page 56, lines 8-10; page 65, lines 1-3). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include wherein determining genotype comprises using smFISH, as previously disclosed by Harvard, in order to provide an improved method for determining the genotype of a cell.

As per claim 159, MSI discloses a method (abstract), comprising: introducing, into a plurality of cells, a nucleic acid comprising an identification portion and an expression portion (transfecting an appropriate eukaryotic host cell, a test DNA sequence comprising a test promoter coupled with a reporter; abstract); and determining phenotype of the plurality of cells (determining the phenotype by FACS; page 10, lines 13-18). MSI does not disclose by sequentially exposing the plurality of cells to nucleic acid probes, and determining binding of the target sequences of the nucleic acid probes within the plurality of cells. Harvard discloses sequentially exposing the plurality of cells to nucleic acid probes (sequentially applying a plurality of nucleic acid probes to a cell sample; abstract; page 14, lines 22-26), and determining binding of the target sequences of the nucleic acid probes within the plurality of cells (the binding of the probes with the target sequences within the sample are determined using fluorescence, which identifies the locations of the nucleic acid probes within the sample; abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of MSI, to include by sequentially exposing the plurality of cells to nucleic acid probes, and determining binding of the target sequences of the nucleic acid probes within the plurality of cells, as previously disclosed by Harvard, in order to provide an improved method for identifying a cell based on phenotype.

Claim 149 lacks an inventive step under PCT Article 33(3) as being obvious over Mirkin in view of US 2008/0269068 A1 to Church et al. (hereinafter 'Church').

As per claim 149, Mirkin discloses a composition (paragraph [0004]; claim 1), comprising: a plurality of nucleic acids (a plurality of nucleic acids; paragraph [0125]; claim 1) comprising an identification portion and an expression portion (comprising a DNA barcode (an identification portion) and a specific binding complement to a target analyte (an expression portion); claim 1), wherein the identification portion comprises N variable portions, N being at least 2 (wherein the DNA barcode has a sequence having at least two portions; claim 1), each variable portion being of at least two possibilities (the DNA barcode has sequences that are different and that serve as an identifier for a particular target analyte; claim 1), and wherein each possible combination of variable portions is present within the plurality of nucleic acids (each type of particle complex probe contains a barcode for a particular target analyte; paragraph [0115]). Mirkin does not disclose wherein the identification portion comprises N variable portions, N being at least 3. Church discloses a barcode comprising an identification portion (paragraph [0063]), wherein the identification portion comprises N variable portions, N being at least 3 (each barcode has three main parts: 1) a degenerate 'query' portion which base pairs with the unknown region of the template; 2) a fixed portion which directs the degenerate 'query' portion to the unknown region by base pairing with fixed sequence on each side of the unknown tag, and 3) barcode which correlates with the identity of one base in the degenerate 'query' portion and is interrogated by hybridization once on the instrument; paragraph [0063]). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include wherein the identification portion comprises N variable portions, N being at least 3, as previously disclosed by Church, in order to provide a superior barcode for use in the composition.

-***-Continued Within the Next Supplemental Box-***-

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:

-***-Continued from Previous Supplemental Box-***-

Claims 163 and 164/162-163 lack an inventive step under PCT Article 33(3) as being obvious over Mirkin in view of MSI.

As per claim 163, Mirkin discloses the method of claim 162, but Mirkin does not disclose randomly pairing an identification portion and an expression portion to form a nucleic acid. MSI discloses randomly pairing an identification portion and an expression portion to form a nucleic acid (fusing random promoters to GFP proteins for FACS identification; page 10, lines 19-21). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include randomly pairing an identification portion and an expression portion to form a nucleic acid, as previously disclosed by MSI, in order to provide an improved library of nucleic acids that exhibits random identification portions and an expression portions.

As per claim 164/162, Mirkin discloses the method of claim 162, but Mirkin does not disclose determining pairing between expression portion and identification. MSI discloses determining pairing between expression portion and identification (identifying the test promoter and reporter combinations; page 23, lines 12-16). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include determining pairing between expression portion and identification, as previously disclosed by MSI, in order to provide an improved library of nucleic acids that exhibits random identification portions and an expression portions.

As per claim 164/163, Mirkin and MSI, in combination, disclose the method of claim 163, but Mirkin does not disclose determining pairing between expression portion and identification. MSI discloses determining pairing between expression portion and identification (identifying the test promoter and reporter combinations; page 23, lines 12-16). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Mirkin, to include determining pairing between expression portion and identification, as previously disclosed by MSI, in order to provide an improved library of nucleic acids that exhibits random identification portions and an expression portions.

Claims 3 and 4/3 lack an inventive step under PCT Article 33(3) as being obvious over Goldbard in view of MSI, and in further view of US 2004/0137488 A1 to Kenten et al. (hereinafter 'Kenten').

As per claim 3, Goldbard and MSI, in combination, disclose the method of claim 2, but Goldbard does not disclose wherein the transfected cells are selected using antibiotic selection. Kenten discloses a method of transfecting a cell with a gene of interest and detecting changes in the phenotype of the cell (paragraph [0018]), wherein the transfected cells are selected using antibiotic selection (wherein the transfected cells are selected using antibiotic selection; paragraph [0037])

As per claim 4/3, Goldbard, MSI, and Kenten, in combination, disclose the method of claim 3, and Goldbard further discloses wherein the transfected cells are sorted (labeled cells are detected by FACS analysis; column 11, lines 33-34). Goldbard does not disclose transfected cells. MSI discloses transfected cells (abstract). It would have been obvious to a person of ordinary skill in the art, at the time the invention was made, to have modified the disclosure of Goldbard, to include transfected cells, as previously disclosed by MSI, in order to provide a superior cell that has been transfected for easy identification.

Claims 1-3, 4/2-3, 142-143, 144/142-143, 149-153, 154/152-153, 159, 161-163, 164/162-163, and 180-181 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

PCT Recordation of Search History

Case/PCT Application Number: PCT/US18/34651

CLIN Number/Technical Field of PCT Application: 0007 Biotechnology

Date(s) During Which the Search was Conducted: 30 August 2018

Date of Completion of Recordation of Search History Form: 18 September 2018

Research Analyst Initials: CJF

Search Approval Official (SAO) Initials: DGF

Field of Search/Classification Information:

IPC(8): C40B 30/04; C12Q 1/68; G06F 19/20, 19/22, 19/24 (2018.01)

CPC: C12N 15/1065, 15/1079; C12Q 1/68, 1/6813, 1/6841; G06F 19/20, 19/22, 19/24

Databases Searched (Patent and Non-Patent Literature (NPL), Including Sub-Databases and Files Searched) and Search Terms Used:

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); EBSCO Discovery; PubMed; Google Scholar;

KEYWORDS: determine, genotype, phenotype, nucleic acid, identifier

Database Search String Recordation, Including Dates of Searches:

Patent Database Search Strategy/Results:

PatSeer

ID	Search String	Records	Created Date
L25	TACD: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*)AND ((ASN: "PRESIDENT AND FELLOWS OF HARVARD COLLEGE") OR (INV: ZHUANG WD0 Xiaowei) OR (INV: EMANUEL WD0 George) OR (INV: MOFFITT WD0 Jeffrey))	0	30-Aug-18
L24	TAC: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*)AND ((ASN: "PRESIDENT AND FELLOWS OF HARVARD COLLEGE") OR (INV: ZHUANG WD0 Xiaowei) OR (INV: EMANUEL WD0 George) OR (INV: MOFFITT WD0 Jeffrey))	0	30-Aug-18
L23	TA: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*)AND ((ASN: "PRESIDENT AND FELLOWS OF HARVARD COLLEGE") OR (INV: ZHUANG WD0 Xiaowei) OR (INV: EMANUEL WD0 George) OR (INV: MOFFITT WD0 Jeffrey))	0	30-Aug-18
L22	L19 AND L15 AND ((IC: C40B30/04 OR C12Q1/68 OR G06F19/20 OR G06F19/22 OR G06F19/24) OR (CPC: C12N15/1065 OR C12N15/1079 OR C12Q1/68 OR C12Q1/6813 OR C12Q1/6841 OR G06F19/20 OR G06F19/22 OR G06F19/24))	63	30-Aug-18
L21	TACD: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*) AND ((IC: C40B30/04 OR C12Q1/68 OR G06F19/20 OR G06F19/22 OR G06F19/24) OR (CPC: C12N15/1065 OR C12N15/1079 OR C12Q1/68 OR C12Q1/6813 OR C12Q1/6841 OR G06F19/20 OR G06F19/22 OR G06F19/24))	114	30-Aug-18
L20	L19 AND L15	114	30-Aug-18
L19	TA: (determin* OR detect* OR identif*) AND (phenotyp* OR genotype*)	20430	30-Aug-18
L18	L17 AND L15	0	30-Aug-18
L17	TA: (determin* OR detect* OR identif*) AND phenotyp* AND genotype*	871	30-Aug-18
L16	L13 AND L15	0	30-Aug-18
L15	TACD: transfect* W10 ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W10 (tag* OR identif*)	16721	30-Aug-18
L14	L13 AND L12	0	30-Aug-18
L13	TA: (determin* OR detect* OR identif*) W10 phenotyp* W10 genotype*	212	30-Aug-18
L12	TACD: (determin* OR detect* OR identif*) W20 phenotyp* W20 genotype* AND transfect* W10 ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W10 (tag* OR identif*)	953	30-Aug-18
L11	TAC: (determin* OR detect* OR identif*) W20 phenotyp* W20 genotype* AND transfect* W10 ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W10 (tag* OR identif*)	2	30-Aug-18

L10	TA: (determin* OR detect* OR identif*) W20 phenotyp* W20 genotype* AND transfect* W10 ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W10 (tag* OR identif*)	0	30-Aug-18
L9	TACD: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*)	306	30-Aug-18
L8	TAC: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*)	0	30-Aug-18
L7	TA: (determin* OR detect* OR identif*) W50 phenotyp* AND genotype* AND transfect* ((nucleic* W6 acid*) OR polynucleotide* OR DNA* OR RNA*) W20 (tag* OR identif*)	0	30-Aug-18
L6	TACD: (determin* OR detect* OR identif*) W50 phenotyp* W50 cell* AND introdu* W50 (tag* OR identif*) W50 express* AND (determin* OR detect* OR identif*) W50 genotyp* W50 prob* bind* W50 target*	0	30-Aug-18
L5	TAC: (determin* OR detect* OR identif*) W50 phenotyp* W50 cell* AND introdu* W50 (tag* OR identif*) W50 express* AND (determin* OR detect* OR identif*) W50 genotyp* W50 prob* bind* W50 target*	0	30-Aug-18
L4	TA: (determin* OR detect* OR identif*) W50 phenotyp* W50 cell* AND introdu* W50 (tag* OR identif*) W50 express* AND (determin* OR detect* OR identif*) W50 genotyp* W50 prob* bind* W50 target*	0	30-Aug-18
L3	TACD: (determin* OR detect* OR identif*) W20 phenotyp* W20 cell* AND introdu* W20 (tag* OR identif*) W20 express* AND (determin* OR detect* OR identif*) W20 genotyp* W20 prob* bind* W20 target*	0	30-Aug-18
L2	TAC: (determin* OR detect* OR identif*) W20 phenotyp* W20 cell* AND introdu* W20 (tag* OR identif*) W20 express* AND (determin* OR detect* OR identif*) W20 genotyp* W20 prob* bind* W20 target*	0	30-Aug-18
L1	TA: (determin* OR detect* OR identif*) W20 phenotyp* W20 cell* AND introdu* W20 (tag* OR identif*) W20 express* AND (determin* OR detect* OR identif*) W20 genotyp* W20 prob* bind* W20 target*	0	30-Aug-18

Non-Patent Literature (NPL) Search Strategy/Results:

Google Scholar: 30 August 2018

Search String	Results
determine genotype cells transfect nucleic acid identification portion expression portion	21100
determine phenotype cells tag introduce expression portion determine genotype nucleic acid probes bind target	19000
determine genotype sequentially expose cells nucleic acid probes determine binding target sequences	17300

PubMed: 30 August 2018

Search String	Results
determine genotype cells transfect nucleic acid identification portion expression portion	0

determine phenotype cells tag introduce expression portion determine genotype nucleic acid probes bind target	0
determine genotype sequentially expose cells nucleic acid probes determine binding target sequences	0

EBSCO Discovery: 30 August 2018

Search String	Results
determine genotype cells transfect nucleic acid identification portion expression portion	2155
determine phenotype cells tag introduce expression portion determine genotype nucleic acid probes bind target	11203
determine genotype sequentially expose cells nucleic acid probes determine binding target sequences	8635

IFD

Wolf, Greenfield & Sacks, P.C.

FEB 07 2018

Docketed ✓ Already Docketed
Not Required
Initials 1st 2nd

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 05FEB 2018
Applicant's or agent's file reference H0498.70645	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 17/60558	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee W. Young PCT Helpdesk: 571-272-4300 Telephone No. PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2014)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)*
Applicant's or agent's file reference H0498.70645	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 17/60558	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   How? Directly to the International Bureau of WIPO preferably through ePCT or on paper to, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 338 82 70

   For more detailed instructions, see *PCT Applicant's Guide*, International Phase, paragraphs 9.004 – 9.011.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to any protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:
   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with any request to forward the texts of both the protest and the decision thereon to the designated Offices.
   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   The applicant may **submit comments on an informal basis on the written opinion of the International Searching Authority** to the International Bureau. These comments will be made available to the public after international publication. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established.

   Shortly after the expiration of **18 months from the priority date, the international application will be published** by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau before the completion of the technical preparations for international publication (Rules 90*bis*.1 and 90*bis*.3).

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for **entry into the national phase** before those designated Offices. In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months. For details about the applicable time limits, Office by Office, see www.wipo.int/pct/en/texts/time_limits.html and the *PCT Applicant's Guide*, National Chapters.

   Within **22 months from the priority date, the applicant may request that a supplementary international search be carried out** by a different International Searching Authority that offers this service (Rule 45*bis*.1). The procedure for requesting supplementary international search is described in the *PCT Applicant's Guide*, International Phase, paragraphs 8.006-8.032.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee W. Young PCT Helpdesk: 571-272-4300 Telephone No. PCT OSP: 571-272-7774

Form PCT/ISA/220 (July 2014)

RECEIVED
Wolf, Greenfield & Sacks, P.C.
FEB 07 2018
Docketed ___ Already Docketed ___
Not Required ___
Initials 1st ___ 2nd ___

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference H0498.70645	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 17/60558	International filing date *(day/month/year)* 08 November 2017 (08.11.2017)	(Earliest) Priority Date *(day/month/year)* 08 November 2016 (08.11.2016)
Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☒ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 2A

☒ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

Form PCT/ISA/210 (first sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60558

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-57
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/60558

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G01N 1/36, C12Q 1/68 (2017.01)
CPC - C12Q 1/6841, C12Q 1/6806, G01N 2001/364, G01N 1/36

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/0304952 A1 (Massachusetts Institute Of Technology et al.) 20 October 2016 (20.10.2016) para [0015]-[0022]; [0027]; [0029]; [0037]; [0049]	1-4, 58-61
X	Chen et al. "Nanoscale imaging of RNA with expansion microscopy" Nature Methods, 2016; 13:679-684. Published online 04 July 2016 (04.07.2016) (doi: 10.1038/nmeth.3899) whole document	1-4, 58-61
A	US 2016/0116384 A1 (Massachusetts Institute Of Technology) 28 April 2016 (28.04.2016) whole document	1-4, 58-61
A	Tillberg et al. "Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies" Nat Biotechnol. 2016; 34(9):987-992. Published online 04 July 2016 (04.07.2016) (doi: 10.1038/nbt.3625.) whole document	1-4, 58-61
A	Chen et al. "Expansion microscopy" Science, 30 Jan 2015 (30.01.2015): 347(6221):543-548 (doi: 10.1126/science.1260088) whole document	1-4, 58-61

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 January 2018	05 FEB 2018

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

RECEIVED
Wolf, Greenfield & Sacks, P.C.
FEB 07 2018
Docketed Already Docketed
Not Required
Initials 1st 2nd

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To: TANI CHEN
WOLF, GREENFIELD & SACKS, P.C.
600 ATLANTIC AVENUE
BOSTON, MA 02210-2206

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	05FEB 2018

Applicant's or agent's file reference	FOR FURTHER ACTION
H0498.70645	See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 17/60558	08 November 2017 (08.11.2017)	08 November 2016 (08.11.2016)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G01N 1/36, C12Q 1/68 (2017.01)
CPC - C12Q 1/6841, C12Q 1/6806, G01N 2001/364, G01N 1/36

Applicant PRESIDENT AND FELLOWS OF HARVARD COLLEGE

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	04 January 2018	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (January 2015)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

a. ☐ forming part of the international application as filed:

☐ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 5-57

because:

☐ the said international application, or the said claims Nos. _____ relate to the following subject matter which does not require an international search *(specify)*:

☒ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. 5-57 are so unclear that no meaningful opinion could be formed *(specify)*:

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for said claims Nos. 5-57

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in the form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13*ter*.1(a) or (b).

☐ See Supplemental Box for further details.

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	59	YES
	Claims	1-4, 58, 60-61	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-4, 58-61	NO
Industrial applicability (IA)	Claims	1-4, 58-61	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-4, 58, 60-61 lack novelty under PCT Article 33(2) as being anticipated by US 2016/0304952 A1 to Massachusetts Institute Of Technology et al. (hereafter, MIT).

Regarding claim 1, MIT teaches a method, comprising: embedding cells within an expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material; para [0027]; As used herein, the term "biological sample" means any biological sample that comprises, or is believed to comprise, nucleic acid sequences including, but not limited to, cDNA, mRNA and genomic DNA. Such samples may include tissues containing a variety of cells, a single cell or organelles. Suitable biological samples include, but are not limited to, a blood sample, biopsy specimen, tissue explant, organ culture, biological fluid or any other tissue or cell preparation, or fraction or derivative thereof or isolated therefrom); immobilizing nucleic acids from the cells to the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material); expanding the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample); exposing the expandable material to a plurality of nucleic acid probes (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... (g) sequencing the nucleic acids present in the first enlarged sample; para [0032]; General sequencing methods known in the art, such as sequencing by ... fluorescent in situ sequencing (FISSEQ); para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction); and determining binding of the nucleic acid probes to the immobilized nucleic acids (para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction; washing away unincorporated nucleotide, detecting incorporation of the nucleotide by measuring fluorescence, and repeating the cycle).

Regarding claim 2, MIT teaches the method of claim 1, wherein immobilizing the nucleic acids from the cells to the expandable material comprises immobilizing the nucleic acids at a single point to the expandable material (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that the RNA samples are only immobilized at one point which is the 3' end)).

Regarding claim 3, MIT teaches the method of claim 1, comprising immobilizing the nucleic acids to the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material).

Regarding claim 4, MIT teaches the method of claim 3, comprising immobilizing the nucleic acids at a 5' end of the nucleic acids to the expandable material (para [0029]; acrylamide modified oligonucleotide primers may be covalently fixed within a swellable material such as a polyacrylate gel. As used herein, the term "acrylamide modified" in reference to an oligonucleotide means that the oligonucleotide has an acrylamide moiety attached to the 5' end of the molecule).

--- V.2 Citations and explanations ---

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 17/60558

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
--- Continued in next Supplemental Box ---

Regarding claim 58, MIT teaches a method, comprising: immobilizing a plurality of targets to an expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material), wherein at least 50% of the plurality of targets immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)); and expanding the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample).

Regarding claim 60, MIT teaches an article, comprising: an expandable material (para [0037]; Preferably, the swellable material is a swellable polymer or hydrogel), comprising an embedded cell and a plurality of nucleic acids immobilized to the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material), wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)).

Regarding claim 61, MIT teaches an article, comprising: a polymer comprising an expanded cell and a plurality of nucleic acids immobilized to the polymer (para [0037]; Preferably, the swellable material is a swellable polymer or hydrogel; para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample), wherein the cell is expanded to at least 5 times its normal size within the polymer (para [0049]; In one embodiment, the addition of water allows for the embedded sample to expand at least 3 times, preferably 4 times, preferably 5 times, or more its original size in three-dimensions), and wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)).

Claim 59 lacks an inventive step under PCT Article 33(3) as being obvious over MIT.

Regarding claim 59, MIT teaches a method, comprising: immobilizing a plurality of nucleic acids to an expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: a) attaching target nucleic acids present in the biological sample with a small molecule linker or nucleic acid adapter; b) embedding the biological sample in a swellable material wherein the small molecule linker or nucleic acid adaptor is attached both to the target nucleic acids present in the sample and to the swellable material), wherein at least 50% of the plurality of nucleic acids immobilized to the expandable material are immobilized at single points (para [0029]; anchorable DNA adaptors may be ligated to the 3' ends of the RNAs in the sample with RNA ligases; (since the RNA samples are ligated at the 3' ends, it would follow that all the RNA samples are only immobilized at a single point which is the 3' end)); exposing the expandable material to a plurality of nucleic acid probes (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... (g) sequencing the nucleic acids present in the first enlarged sample; para [0032]; General sequencing methods known in the art, such as sequencing by ... fluorescent in situ sequencing (FISSEQ); para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction); expanding the expandable material (para [0015]-[0022]; Preferably, the invention provides a method for in-situ sequencing of target nucleic acids present in a biological sample comprising the steps of: ... d) swelling the swellable material to form a first enlarged biological sample that is enlarged as compared to the biological sample); and determining binding of the nucleic acid probes to the immobilized nucleic acids (para [0033]; FISSEQ is a method whereby DNA is extended by adding a single type of fluorescently-labelled nucleotide triphosphate to the reaction; washing away unincorporated nucleotide, detecting incorporation of the nucleotide by measuring fluorescence, and repeating the cycle);
but does not specifically teach wherein the step of exposing the expandable material to a plurality of nucleic acid probes occurs before expanding the expandable material. Nonetheless, since MIT teaches the steps, it would have been obvious to one of ordinary skill in the art to trivially switch the order of labelling steps and expanding the expandable material during routine experimentation to optimize the method.

Claims 1-4, 58-61 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

SEARCH HISTORY

Application Number	PCT/US 17/60558
Search Conducted By	FE
Search Approved By	SV

CPC/IPC Classifications Searched	IPC (8) – G01N 1/36, C12Q 1/68 (2017.01) CPC – C12Q 1/6841, C12Q 1/6806, G01N 2001/364, G01N 1/36
Date Conducted	4 January 2018 (04.01.2018)

Documentation Searched	IPC (8) – G01N 1/36, C12Q1/68 (2017.01) CPC – C12Q 1/6841, C12Q 1/6806, G01N 2001/364, G01N 1/36
Search Terms Used	embedding cells expandable material immobilizing nucleic acids probes MERFISH ExFISH water swelling polymer cell culture expansion microscopy
Date Conducted	4 January 2018 (04.01.2018)

Electronic Database Searched	PatBase
Files Searched	Full-text: AU BE BR CA CH CN DE DK EP ES FI FR GB IN JP KR SE TH TW US WO Bibliographic: (European) AT BA BE BG CH CS CY CZ DD DK EE ES FI GE GR HR HU IE IS IT LT LU LV MC MD MT NL NO PL PT RO RS SE SI SK SM TR UA YU (Asia) EA GC HK ID IL IN KZ MN MY PH RU SG SU TH TJ TW UZ VN (North America) CA CR CU DO GT HN MX NI PA SV TT (South America) AR BR CL CO EC PE UY (Australasia) AU NZ (Africa) AP DZ EG KE MA MW OA ZA ZM ZW
Date Conducted	4 January 2018 (04.01.2018)

Search Logic:

1) IC=(G01N1/36 or C12Q1/68) (132519)
2) CPC=(C12Q1/6841 or C12Q1/6806 or G01N2001/364 or G01N1/36) (5271)
3) 1 or 2 (132912)
4) FT=(((expand* w10 (matrix or substrate or material or support or culture)) w15 cell*) and (nucleic acid or DNA or RNA)) (7228)
5) 3 and 4 (1522)
6) FT=(((expand* w10 (matrix or substrate or material or support or culture)) w15 cell*) and ((nucleic acid or DNA or RNA) w10 probe)) (1858)
7) 6 and 3 (953)

8) FT=((((embed* or grow* or deposit* or attach* or culture*) w10 cell*) w25 expand*) w10 (matrix or substrate or material or support or polymer)) (381)
9) 7 and 8 (3)
10) 6 and 8 (12)
11) FT=((((embed* or grow* or deposit* or attach* or culture* or immobiliz*) w10 cell*) w25 (nucleic acid or DNA or RNA)) w10 probe) (200)
12) 8 and 11 (0)
13) FT=((((single or one) w15 (location or point or spot or position* or anchor*)) w15 (embed* or bound or deposit* or attach* or immobiliz*)) w25 (nucleic acid or DNA or RNA or cell)) (2172)
14) 8 and 13 (3)
15) 11 and 13 (2)
16) FT=(MERFISH or ExFISH) (15)
17) FT=(expansion microscopy) (25)
18) FT=((((expand* or swell*) w10 (matrix or substrate or material or support or culture)) w15 cell*) and (nucleic acid or DNA or RNA) and fluorescen*) (6062)
19) 17 and 18 (2)
20) 16 and 18 (9)

Electronic Database Searched	Google Patents
Files Searched	Google Patents (patents.google.com)
Date Conducted	4 January 2018 (04.01.2018)
Search Logic:	
expandable cell culture material : About 11,400 results (0.55 seconds) expandable cell culture material nucleic acid probe bound : About 3,350 results (0.51 seconds) immobilize cell expandable matrix : About 8,490 results (0.47 seconds) immobilize nucleic acid expandable matrix : About 4,320 results (0.63 seconds) immobilize nucleic acid expandable matrix fluorescent DNA probe : About 1,770 results (0.46 seconds) water swelling expandable cell culture polymer material : About 9,430 results (0.43 seconds) Expansion microscopy : About 102,000 results (0.45 seconds) Expansion microscopy fluorescent DNA probe : About 92,600 results (0.41 seconds) Expansion microscopy fluorescent DNA probe anchor single point five prime end : About 4,690 results (0.49 seconds)	

Electronic Database Searched	Google Scholar
Files Searched	Google Scholar (excluding patents) (scholar.google.com)
Date Conducted	4 January 2018 (04.01.2018)
Search Logic:	
RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH) : About 67 results (0.06 sec) water swelling expandable cell culture polymer material : About 3,830 results (0.08 sec) Expansion microscopy : About 2,050,000 results (0.07 sec)	

Electronic Database Searched	Google
Files Searched	Google Web (www.google.com)
Date Conducted	4 January 2018 (04.01.2018)
Search Logic:	
expandable cell culture material : About 211,000 results (0.55 seconds) expandable cell culture material nucleic acid probe bound : About 175,000 results (1.02 seconds) immobilize nucleic acid expandable matrix : About 212,000 results (0.86 seconds) immobilize nucleic acid expandable matrix fluorescent DNA probe : About 108,000 results (0.82 seconds) RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH) : About 1,120 results (0.62 seconds) water swelling expandable cell culture polymer material : About 152,000 results (1.05 seconds) Expansion microscopy : About 46,200,000 results (0.54 seconds)	

Supplemental

Electronic Database Searched	Google Patents
Files Searched	
Date Conducted	5 January 2018 (05.01.2018)

embedding cells" expandable material "nucleic acid probe"
About 286 results (0.45 seconds)

"embedding cells" expandable material
About 18 results (0.38 seconds)

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12N 15/10	A1	(11) International Publication Number: WO 99/31277
		(43) International Publication Date: 24 June 1999 (24.06.99)

(21) International Application Number: PCT/US98/26715

(22) International Filing Date: 15 December 1998 (15.12.98)

(30) Priority Data:
9726431.1 15 December 1997 (15.12.97) GB

(71) Applicant *(for all designated States except US)*: MEDICAL SCIENCE SYSTEMS, INC. [US/US]; 3567 Shaw Avenue, Cincinnati, OH 45208 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DOWER, Steven [GB/GB]; University of Sheffield – The Medical School, Division of Molecular and Genetic Medicine, Royal Hallamshire Hospital, Sheffield S10 2JF (GB). QUARNSTROM, Eva [GB/GB]; University of Sheffield – The Medical School, Division of Molecular and Genetic Medicine, Royal Hallamshire Hospital, Sheffield S10 2JF (GB).

(74) Agents: ARNOLD, Beth, E. et al.; Foley, Hoag & Eliot, LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: CA, IL, IN, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: EXPRESSION CLONING AND SINGLE CELL DETECTION OF PHENOTYPE

(57) Abstract

The present invention pertains to a new method of "functional cloning" or "expression cloning" based on single–cell detection of a conferred phenotype. More particularly, in a preferred embodiment the method involves co–transfecting an appropriate eukaryotic host cell, a test DNA sequence and an appropriate CDNA library. The test DNA sequence contains a test promoter coupled with a reporter. CDNA's coding for molecules that influence the expression of the reporter could be detected.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

EXPRESSION CLONING AND SINGLE CELL DETECTION OF PHENOTYPE

BACKGROUND

1. Field of the Invention

This invention pertains to a new method of "functional cloning" or "expression cloning" based on single-cell detection of a conferred phenotype. More particularly, in a preferred embodiment the method involves co-transfecting an appropriate eukaryotic host cell a test DNA sequence and an appropriate cDNA library, both of eukaryotic origin. The test DNA sequence contains a test promoter or portion thereof coupled with a reporter. The preferred reporter encodes the green fluorescent protein (GFP) or one of its variants. The cDNA library is preferably constructed in a high expression vector. If a cDNA present in the library codes for a molecule that influences the expression of the reporter, this will be seen as a change in fluorescence and can be detected in a single cell. The cell can thus be identified and the cDNA from it recovered in a number of ways.

The preferred method of identifying and selecting positive cells is fluorescence activated cell sorting (FACS). The cDNAs from an enriched population of positive cells can be purified and retransformed into bacterial cells. Positive cDNA clones can be tested by repeating the co-transfection experiment and analyzing the co-transfected cells by fluorescence microscopy. The procedure has been exemplified with a portion of the IL-8 promoter and it has been demonstrated that the pIL-8 construct is sensitive to cytokine activation in this co-transfection system.

2. Description of the Prior Art

"Chronic inflammatory diseases" are a wide spread health problem in the western world. This broad term covers a spectrum of conditions that include arthritis, atherosclerosis, various dermatological disorders such as psoriasis and other specific inflammatory diseases such as inflammatory bowel disease. It is a

broadly held view that the underlying driving mechanism of inflammatory diseases is cytokine over-expression. Cytokines produced transiently in normal responses to injury and infection can become persistently produced, leading to chronic inflamation. Such diseases tend to have a slow course, often spanning many years before symptoms become clinically significant. Consequently, the causes of the failure of the initial inflammatory responses are obscure, but likely derive from a complex interplay of environmental and genetic factors.

Current methods for treating such diseases employ surgical, immuno-suppressive (e.g., total lymphoid irradiation), and non-specific anti-inflammatory treatments (e.g., steroids). "Anti-cytokine" therapies based on inhibition of receptor/ligand binding for particular cytokines, while promising, have met with limited clinical success. Therefore, the development of a deeper understanding of the mechanisms of the gene regulation and action of pro-inflammatory cytokines continues to attract a great deal of interest in many academic and industrial institutions. Illucidation of the signal transduction pathways involved in cytokine activation of genes and identification of the components of these pathways and their model of action is therefore of importance.

Functional cloning methods have increasingly replaced conventional biochemical approaches to identifying genes of interest (Dower, S. (1993) *in Neurobiology of Cytokines,* vol. 16., E.B. DeSouza, ed. Academic Press, Inc., New York, p. 3-30). A particularly striking example is the application of the yeast two hybrid system for identifying partners to intracellular proteins in mammalian cells based on pairwise binding interactions which reconstitute transcription systems in yeast (Fields, S. and Song, 0. (1989) *Nature* 340:245-6).

Based on the same principle (pairwise protein-protein binding), mammalian high expression cloning systems have proved particularly effective at identifying receptors for soluble factors and ligands for orphan receptors (Seed, B.

(1995) *Current Opinion In Biotechnology* 6:567-573). This method usually relies on the use of autoradiography or immunocytochemistry to identify the presence of a clone in a complex pool of cDNAs that encodes the partner for a labeled probe molecule. Operationally, a library of cDNAs (typically 1-5 x 10^6 independent transformants), prepared using random hexanucleotide primers (to minimize the representation of intrinsic 3' non coding regions), is cloned into a mammalian expression vector, which contains at least an *E. coli* origin, a selectable marker, and a high expression cassette (e.g., constaining a strong promoter such as CMV or adenovirus MLP; a polylinker and a splice/viral 3' noncoding/polyA+ tail). The library is transformed into *E. coli,* plated and pools of plasmids from 1,000 - 10,000 independent colonies are prepared.

Each pool is transfected into an indicator line, such as COS-7, and assayed for the phenotype of interest. Thus, when searching for a receptor, radio-labeled ligand is added, and cells binding high levels of label are detected by autoradiography. Typically, the presence of an appropriate cDNA in a pool of 10,000 will be indicated by 20-100 positive cells in a monolayer on a microscope slide of about 2 X 10^5 COS-7 cells. The positive pool can then be broken down by replating an aliquot of bacteria held in reserve against the identification of that pool, and repeating the cycle. Three to four cycles plus a round of single colony picking, a process that takes about 1 month, will usually suffice to isolate a cDNA of interest.

This method is disadvantaged, however, in that single positive cells cannot be isolated and identified, requiring the sequential dilution and screening of pools of clones in order to narrow the search until a single positive clone is discovered. Furthermore, the currently used methds do not allow for the detection and isolation of single mammalian cells containing a test cDNA of interest. What is needed in the art, is a selection system that can be generalized to rapidly detect a

phenotype conferred on mammalian cells at the single cell level. The invention described herein is such a cloning system, allowing single cell detection and isolation of positive cells.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Structure of the pIL-8 Construct

Figure 2: Fluorescent Micrographs Converted to Histogram Format

Figure 3: Effect of IL-1RI Transfection and IL-1 Stimulation on the Activity of Reporter Constructs in Cos-7 Cells

SUMMARY OF THE INVENTION

Briefly, the invention is a method of isolating a nucleotide sequence that increases (or otherwise modulates) expression from a test promoter. The method involves transfecting a mammalian host cell with a first nucleotide sequence to produce a transfected cell, wherein the first nucleotide sequence has at least a test promoter coupled with a reporter. Although any appropriate cell/promoter/reporter combination can be used, preferably, the mammalian host cell is a COS or Hela cell, the test promoter is a cytokine promoter, such as the IL-8 promoter or portion thereof, and the reporter encodes a fluorescent protein, such as green fluorescent protein (GFP).

Next, the transfected cell is co-transfected with a second nucleotide sequence to produce a co-transfected cell, wherein the second nucleotide sequence has at least a cDNA coupled to a high expression promoter. The expression of the reporter in the co-transfected cell is measured and compared with a control cell transfected only with the first nucleotide sequence.

In an alternate embodiment of the invention, one or more agents, such as, for example, a cytokine or other hormone, may be provided to the cell to

determine if the reporter expression is further increased in response to the agent(s). The agent may be anything that regulated the expression of a gene and may include also chemical agents, inhibitory agents such as antisense molecules, peptides, metal, depending upon the signaling pathway that is being illucidated.

Alternatively, an additional DNA sequence may be transfected into the cells to further enhance the synthesis of the reporter protein. For example, a cDNA encoding a receptor for the agent that is being provided to the cell as described above may be desired. In this way the cell may produce additional receptors with which the agent may interact, thereby increasing the level of signaling molecules within the cell. For example, the preferred embodiment of the invention may be altered to included the co-transfection of a cDNA encoding the IL-1 receptor (IL-1R). The resultant cells containing the IL-8-promoter-GFP cDNA and the IL-1R cDNA, thereby expressing increased levels of IL-1 R. These cells will thereby be more responsive to IL-1 treatment. Alternatively, other signaling molecules may be introduced, including, traf-6, myd-88, FADD, NIK, IRAK1, IRAK2, and REIA.

Although in the experiments devised herein, we have selected as positive those cDNA clones that increase the expression of a test promoter, it is of course equally possible to select for clones that decrease the expression of a highly active test promoter.

In yet another embodiment of the invention, the method allows for the enrichment of a cDNA library for a cDNA that modulates the expression of a test promoter. The method involves transfecting mammalian host cells with a first nucleotide sequence (as described above) to create transfected cells and then co-transfecting the transfected cells with a library of cDNA molecules to create a co-transfection library. Preferably, each cDNA molecule is coupled to a high expression promoter. Next, the co-transfection library is sorted by a fluorescence

activated cell sorter to create an enriched co-transfection cell library, wherein the co-transfection cells that produce a changed level of fluorescence, as compared with transfection control cells, are positively selected for inclusion in the enriched co-transfection cell library. Finally, the cDNA molecules from the enriched co-transfection library are purified, re-transformed into bacterial cells, amplified and repurified to create an enriched library of cDNA molecules. The method can be repeated until a sufficient level of enrichment for cDNA molecules that modulate the expression of the test promoter is achieved and each cDNA molecule can be isolated.

DETAILED DESCRIPTION OF THE INVENTION

Abbreviations:

BSA: bovine serum albumin
CMV: cytomegalovirus
FACS: fluorescence activated cell sorter
GFP: green fluorescent protein
IL-1 interleukin-1
IL-8: interleukin-8
IL-1R: interleukin-1 receptor
IL-1RI: interleukin-1 receptor type I
PBS: phosphate buffered saline
TNF: tissue necrosis factor

Definitions:

As used herein, the term "GFP" refers to green fluorescent protein or any mutant or variant thereof.

As used herein, the term, "gene regulatory sequences" refers to DNA or RNA sequences that act in cis to modulate gene transcription or a post-transcriptional mechanism that regulates protein synthesis. Such sequences can include, but are not limited to promoter sequences, enhancer sequences, other regions of DNA required for gene expression, and regions of RNA molecules that regulate, for example, mRNA turnover, translation efficieny, or polyadenylation.

As used herein, the term "reporter" or "reporter protein" refers to a nucleotide sequence encoding a detectable product (e.g., a protein). The reporter protein may be fluorescently, radioactively, biologically and/or enzymatically detectable. For example, the reporter protein may be detected by fluorescence activated cell sorting or fluorescence microscopy. Preferred reporter proteins are green fluorescent protein (GFP), EGFP™, d2-EGFP, S6ST, T203Y and GFPmut or other GFP variant. Alternative fluorescent molecules may be used which differ from GFP in terms of its excitation wavelength (e.g., is a different color such as blue (BFP), yellow, orange or red) or its fluorescent intensity. Any reporter protein that is which is conducive to the use of the invention with live cells is preferred.

As used herein, the term "test promoter" or "test promoter sequence" refers to a promoter, or portion thereof, for which it is desired to identify and clone sequences that influence its expression. In this application, the IL-8 test promoter is exemplified, but any promoter which is a component of a signal transduction pathway make be analysed by the method of the invention.

As used herein, the term "test nucleotide sequence" or "test nucleic acid" refers to a nucleic acid which encodes a factor or polypeptide capable of interacting with a test promoter or other gene regulatory sequence. Preferred

nucleotide sequences encode a transactivating factor, such as, for example, a transcription factor or enhancer molecule.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer.

In one aspect, the instant invention relates to a novel cloning method for identifying transactivating factors that participate in gene regulation. In one embodiment, the invention relates to the identification of transcription factors or other transactivating regulatory factor required for the transcription of a gene. In an alternative embodiment, the invention relates to the identification of *cis*-acting regions of genes that interact with the transactivating factors of the invention. In another embodiment, the invention relates to the determination of interactions between a test transactivating factor and the *cis*-acting region of a gene. In yet another embodiment, the invention relates to the above described embodiments which additionally involve the step of presenting to the transfected cells of the invention an agent which may modulate (e.g., enhance or inhibit) the ability of the transactivating factor to activate the *cis*-acting region of the gene.

In a preferred embodiment, the instant invention relates to a cell signaling pathway dissection procedure, wherein the components of the pathway, and their role in signal transduction, are determined using the present method. For example, cellular signaling components that lie on the pathway that functionally connects a hormone receptor to a gene promoter can be identified and characterized. The method of the invention can be used to clone any nucleotide sequence whose product, or which itself, modulates the expression of a test promoter. It is also anticipated that other regulatory sequences, such as, for example, those present in the 5' untranslated region (UTR) or 3' UTR of a test

mRNA may be analyzed in accordance with the present method. In this embodiment, the method will identify transactivating factors that interact with mRNA molecules to regulate protein synthesis post-transcriptionally. Such regulatory factors include, for example, those that regulate mRNA stability, mRNA polyadenylation or translation efficiency. Adaptation of the instant invention to achieve the study of post-transcriptional gene regulatory pathways would thus require appropriate reporter vectors that contain 5' UTR and 3'UTR sequences of interest.

In a preferred embodiment, the pathways functionally linking IL-1 and TNF receptors to the promoter for the IL-8 gene may be identified, as an example of the broadly applicable novel cloning and pathway dissection procedure of the invention. IL-8 is a down-stream mediator of IL-1 and TNF action, playing a major role in activating the binding of neutrophils to, and their migration across, the endothelium of inflamed tissue (Huber, A., *et al.* (1991) *Science* 254:99-102). As in other IL-1 and TNF responsive genes, the IL-8 promoter contains both an AP-1 and an NF-Kβ consensus site, and both have been shown to be critical for function (Sims, J., *et al.* (1994) *in Signal Transduction Through Growth Factor Receptors.*, vol. 3. JIA Press, Greenwich, Conn., USA, London, U.K., p. 197-222). We have previously shown that a small (200 bp) fragment from this promoter confers a high degree of IL-1 inducibility on reporter constructs incorporating it, using transient expression in COS-7 cells (Mitcham, J., *et al.* (1996) *J. Biol. Chem.* 271:5777-5783). The instant invention was used to functionally clone cDNAs coding for signaling components that lie on the pathways connecting the IL-1 and TNF receptors to the IL-8 promoter. However, the method is of general applicability and can be used to clone any sequence that modulates the expression of a test promoter provided the host cell line is appropriately chosen and contains all of the required ancillary components.

In a preferred embodiment, the reporter protein is a fluorescing protein that can be expressed in a eukaryotic cell and not cause damage to the cell. In a particularly preferred embodiment, green fluorescent protein (GFP) from Aequorea victoria (Prasher, D.C. (1992) *Gene 111:229-233*). GFP is strongly fluorescent in GFP transfected *E. coli* or *C. elegans* cells and its fluorescence does not appear to depend upon exogenous substtrates or coenzymes. (Chalfie, N. et al. (1994) *Science* 263:802-805). GFP has also been used to construct chimeric proteins that fluoresce in Drosophila melanogaster (Wang and Hazelrigg (1994) *Nature* 369:400-403) and in human embryonic kidney cells (Marshall, et al. (1995), *Neuron* 14:211-215). More recently, GFP has been used to study *in vivo* gene expression and protein localization in eukaryotic cells (Cubitt, A.B. et al. (1995), *Trends Biochem. Sci. 20:448-455*).

GFP has been employed in a bacterial expression cloning system that differs from that described here (Valdivia, R.H. and Falkow, S. (1996), *Mol. Microbiol.* 22(2):367-78). Valdivia and Falkow applied a novel enrichment strategy, termed differential fluorescence induction (DFI), to screen a *Salmonella typhimurium* library for promoters that are upregulated at pH 4.5. DFI utilizes GFP and a fluorescence activated cell sorter (FACS) to perform genetic selection. In the presence of an inducing stimulus, such as low pH, a FACS is used to sort highly fluorescent bacterial clones bearing random promoters fused to a mutant GFP protein (GFPmut). This population is then amplified at neutral pH and the least fluorescent population is sorted to eliminate constitutive promoters. Sequential sorts for fluorescent and non-fluorescent bacteria in the presence or absence of inducing conditions enriches for promoter fusions that are regulated by the inducing stimulus.

The technique described by Valdivia and Falkow differs from the technique described herein in several respects. First, their system allows for the

identification of bacterial promoters, not the analysis of eukaryotic regulatory elements as in the instant invention. The system of Valdivia and Falkow does not relate to the methods of the instant invention which begins with an isolated promoter and identify factors that interact with it. The Valdivia and Falkow method also does not require co-transfection of two nucleotide sequences into eukaryotic host cells in order to determine the interactions of the product of one nucleotide sequence with a region of the other nucletode sequence. Additionally, Valdivia and Falkow's system can isolate *only* those promoters that respond to simple environmental conditions such as low pH, high temperature, absence of essential nutrients, etc. The invention herein, in contrast, is more complex, allowing for the study of eukaryotic pathways and the selection of *any sequence* that modulates a eukaryotic test promoter. Thus, many different types of gene regulatory factors could be isolated and studied by the invention described herein.

The instant invention relates to a single cell transcription reporter assay based on use of the *Aequoria victorea* green fluorescent protein (GFP) under the control of the human IL-8 promoter (pIL-8) to demonstrate the functional cloning of cDNAs important in the expression of the IL-8 test promoter. There is ample evidence from the recent literature that introduction of previously cloned cDNAs encoding components of cascades, such as MAP kinase cascades, can be read out as increased activity in downstream steps in the signaling pathways. Thus, there is reason to believe that this principle can be applied prospectively to identify cDNAs encoding novel signaling molecules. We have used a fragment of the IL-8 promoter generated from human genomic DNA by PCR (Mitcham, *J., et al.* (1996) *J. Biol. Chem.* 271:5777-5783) and a commercially available CLONTECH™ vector to construct a test promoter/reporter construct. The construct is shown in Figure 1 and is called pIL-8 herein.

GFP and its variants are the preferred detection molecules of the invention. GFP has several advantages over other markers. First, it is relatively harmless to cells and no processing is required to visualize it. GFP is unique among fluorescent biomolecules in being composed solely of a polypeptide chain. The fluorescent form of the protein is generated by an auto-oxidation reaction after folding of the initial translation product. The only requirement for GFP to fluoresce is some amount of oxygen and time. Further, GFP photobleaches less quickly than other fluorescent molecules. In addition, because all the information for expression of GFP is contained in its gene sequence, the power of mutagenic techniques can be brought to bear on the generation and selection of GFPs with improved spectral, folding and stability properties. The EGFP in our construct is such a second generation protein, with improved autocatalytic conversion rate and spectral properties relative to wild type.

Detectable proteins useful in the practice of the invention include but are not limited to GFP, EGFP™ from CLONTECH™, d2-EGFP and other variants of GFP, S65T, T203Y, GFPmut, and other GFP mutants selected on the basis of altered spectral properties.

While other detectable labels could also be used (such as radio-labels, bio-labels such as biotin, histamine and enzymatic labels, fluorescent labels such as fluorescein, luciferase, and the like) the power of this system is that the reporter protein produced on stimulation of the test promoter enables direct selection of positive living cells. Thus, any reporter protein that is compatible with living cells and is conducive to enrichment and separation of positive cells may be used.

Expression of GFP can be detected by techniques known to those skilled in the art, including, but not limited to, fluorescence microscopy, FACS, western blotting, immunodetection, and the like. Other labels are detected as appropriate for the type of label.

Any number of mammalian high-expression vectors may be employed in the present invention, including, but not limited to, pCDM8™, pCDNA, Amp™, and pCR-3™. We have tested these vectors in COS-7 cells for GFP expression using DEAE-dextran transfection. All give intense GFP fluorescence at approximately 40-50% transfection efficiency. However, the pCR-3™ vector is preferred because it provides slightly higher transfection frequencies and expression levels. An preferred vector should include a strong viral promoter, multiple cloning site, viral 3' non-coding region, *E. coli* selectable marker (such as Amp or SupF), mammalian selectable marker (such as G418 or GFP), viral origin of replications (such as OriP or SV40 Ori), and the large T or EBNA-1 sequences as appropriate.

Any number of cells may be used as the host cell, requiring only that the host cell contain all of the necessary ancillary components for signaling from a given nucleotide sequence to the test promoter. Thought not a requirement, in order to ensure that the system contains being studied compatible components, it is preferred that the cDNA library be made from cells that are compatible with the host cell. For example, we have shown that cDNA libraries made from CV-1 cells are compatible with COS cells (both cell types are derived from African green monkey cells), at least in their cytokine signal transduction machinery.

The universal host/vector pairing is a COS cell and a vector with the SV40 origin of replication. Alternatively vectors with the OriP origin of replication can be used with CV-1 cells stably expressing EBNA-1. Vectors are available from INVITROGEN™ to transfect CV-1 cells with EBNA-1. Host cells include, but are not limited to, COS cells, human epithelial cells such as 293 or 293T, and Hela cells, but any cell type capable of transfection can be used. Specialized cells types, for example, that are IL-1 deficient, are also contemplated.

Protocols for producing preparing and transfection DNA into host cells can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989) and other standard laboratory manuals.

Example 1: Pilot Study: Detection of GFP in Single Cells in Response to IL-8 Modulators

We describe herein a method for co-transfecting an IL-8 test promoter-GFP reporter construct with a high expression cDNA library into COS-7 cells, and examination of pools of cDNA for the capacity to confer on single cells the following characteristics:

1) Expression of GFP in the absence of an external stimulus, and/or
2) Enhanced expression of GFP in response to low doses (~1pM) of IL-1.

COS-7 monolayers in 4 well Lab-Tek chamber slides were transfected (DEAE dextran/1μg each plasmid/well) (McMahon, C., et al. (1991) *EMBO J.* 10:2821-2832) with constructs containing no promoter (pO), the IL-8 promoter (pIL-8), the CMV promoter (pCMV) or co-transfected with pIL-8 and the human IL-1 receptor in pCDNA/I (pIL-1RI). Transfected cells were cultured for 48 hours to allow high expression of IL-1RI and then treated with 10nM IL-1 for 4 hours. Confocal Fluorescence Micrographs were taken (not shown) at Excitation 488 nm, 20 mW, 700 volts PMT*, 10X air objective. *Except for pCMV 500 volts: these data were then scaled by a 16X multiplication since calibration data show that 50 volts = 2X increase in observed intensity. Images were analyzed in NIH IMAGE® and the numeric output converted to histograms with EXCEL®.

Figure 2 shows the results of image analysis of confocal micrographs, both at 4 and 12 hours post IL-1 addition. Note that the x-axis in each of the three panels is different. In the absence of a functional promoter, GFP expression was

negligible and was not enhanced by IL-1 treatment. Expression driven by the CMV promoter gave approximately a 200-fold increase in GFP expression over background, and was unaffected by IL-1. Expression from the IL-8 promoter gives a basal level of expression of GFP that is about 4-fold over the pO control. IL-1 treatment (4 hours) induced an increase in GFP expression in 30% of the cells by about 2.5-fold over that seen without IL-1, despite the fact that the basal transcription was ongoing for 24 hours by the time of IL-1 addition.

COS cells express about 500 IL-1 receptors per cell (Sims, J., *et al.* (1988) *Science* 241:585-589). However, the data clearly show that these cells are capable of responding to IL-1 even with this low level of receptor expression and therefore contain all the components necessary to transcribe off pIL-8, albeit at low efficiency. These observations are consistent with our findings that a pIL-8-IL2Rα construct, containing the same promoter fragment, was IL-1 responsive in COS-7 cells (Mitcham, J., *et al.* (1996) *J. Biol. Chem.* 271:5777-5783). In order to test the principle that introducing a DNA for an IL-1RI-pIL-8 pathway element into this system will enhance the IL-1 responsiveness of these cells, we co-transfected the IL-1 receptor type I (IL-1RI) into the pIL-8 containing cells. The use of the receptor in this control experiment is rational, since the system should be most sensitive to rate limiting steps in the pathway such as recptor expression. It is logical that the receptor would be a primary control or rate limiting step, being "switched" from the "off" to the "on" state when some effector molecule binds.

The data in Figure 3 show that when IL-1 receptors are over-expressed, there is a considerable increase in the activity of pIL-8 even in the absence of IL-1. We have observed similar results in previous experiments using the pIL-8-IL2Rα reporter and controls (not shown) have shown that this response cannot be blocked with high concentrations of anti-IL-1R monoclonal antibodies or IL-1

receptor antagonist (Id.). The increase in pIL-8 expression was, therefore, not due to endogenous IL-1 production but to spontaneous IL-1R activation when the receptor was expressed at levels of about 10^5 copies/cell. Hence, we could easily detect those cells expressing the IL-1R cDNA.

Stimulation with IL-1 for 4 hours produced a further increase in GFP expression in the majority of transfected cells. The data showed that at saturating DNA levels (about 1 μg/100,000 cells), about 30-40 cells/field (200 cells total/field-1 $7mm^2$) was detected as strongly positive in either the plus or minus IL-1 added to pIL-8/pIL-1RI co-transfected cultures. Therefore, in each $2cm^2$ well there would be about 5,000 positive cells suggesting that a pool size of 500-1000 is feasible for screening (5-10 positive cells/well). Based on these numbers, there is every reason to suppose that this procedure can form the basis of a practical expression screen based on pool screening techniques.

Example 2. Single Cell Detection: Hirt DNA Purification and Transformations

The power of this functional cloning system is that the individual cells can be selected and the cDNA clones therein can be purified immediately on detection. For example, using a cloning ring coupled with the single cell DNA purification techniques of Hirt, the DNA from a positive cell can be purified, and retransformed into bacteria for sequencing and further analysis.

The Hirt preparation of episomal DNA is generally as follows: to 1-3 X 10^6 transfected cells on a plate add 1ml of buffer A (10 mM Tris, pH 7.5, 10 mM EDTA, 0. 6% SDS) and incubate 3-5 minutes at room temperature. Then add 0.25 ml of 5 M NaCl. Gently scrape the plate to avoid shearing the genomic DNA and transfer the entire solution to a microfuge tube. Leave at 40°C for 4 to 12 hours. Spin at 15,000 rpm for 15 minutes. If a small amount of DNA is expected,

add carrier DNA (1-10 μg salmon sperm DNA). Phenol extract, ethanol precipitate and use the purified DNA to transform *E. coli* cells. This procedure can be scaled down to use on single cells isolated with a cloning ring. However, the amount of DNA recovered is very small. Therefore, an alternative approach (described below) uses a FACS to sort and collect positive cells from which DNA is purified and transformed into bacteria for amplification and further screening.

Example 3. Single Cell Detection: PCR Amplification

Because single cell DNA preparations and transformations are technically difficult, it may be desired to first amplify the cDNA insert of interest. This can be done by using oligonucleotides that are unique to the plasmid just 5' of the cDNA insert. The other primer can be selected to complement the 5' end reporter sequence or the junction sequence. Either single positive cells can be amplified in this way, or pools of FACS selected positive cells can be amplified. The PCR amplified DNA can then be re-ligated into a suitable vector and transformed into bacterial cells for ease of manipulation.

PCR amplification is very well known in the art and the details need not be described herein. However, for reference the reader is referred to the following: for general advice about PCR see: http://www.fermentus.lt/fermentus/catalog/PCR/DNA_amp/Protocol.htm; a variety of specific protocols can be found at http://research.nwfsc.noaa.gov/protocols.html; and a protocol for the amplication of a single mammalian cell is provided by Holding, C. and Monk, M. (1990) *Genet. Res.* 55(2): 120.

PCR amplification is a powerful technique, well able to amplify the DNA from single cells, as described above. However, unless the dNTP pools are well

balanced and the magnesium and polymerase levels optimized, the polymerase can tend toward error-prone replication. An error early in the amplification process could profoundly affect the functionality of a given cDNA. If this problem is suspected for a particular clone, it can be corrected by using the PCR amplified DNA or PCR generated clone to isolate the original cDNA from the starting library. However, careful selection of PCR conditions should minimize this problem.

Example 4. Single Cell Detection: FACS Selection of Positive Cells

A simple and powerful approach for positive cell selection is to FACS enrich for a population of positive cells. The histograms in Figure 2 show that the fluorescence in positive cells is much stronger than background fluorescence. Because the GFP-expressing cells are not destroyed by the detection means employed, the living cells can be selected by FACS (Yamasaki, *et al.* (1988) *Science*, 241:825-28). The cDNA constructs therein are then purified (or their inserts amplified and ligated into an appropriate vector) and retransformed into bacteria to create a pool of positive clones. Positive clones can be confirmed by the co-transfection procedure above, sequenced and further analyzed. Though the preferred method of the invention is to detect an increase in reporter gene expression, decreases may also be detected using the invention. Further, an altered reporter readout could be due to the change in color of the reporter protein, for example.

The FACS sorting procedure is generally as follows: COS cells are co-transfected with pIL-8 and library DNA at 50 μg DNA per 10^7 cells. Co-transfected COS cells are lifted from the plate by treatment with trypsin and EDTA. The cells are resuspended at 10^6 cells/ml in PBS/BSA (0.01%). Positive

cells are separated by sorting on a FACSTAR plus (Ar 488 line excite, 530bp filter detect). At 10-50,000 cells per minute, it will require about 1 to 2 hours to sort the library. The brightest 10% cells are subject to HIRT DNA preparation and transformation into *E. coli*. Plasmids are prepared from the transformed *E. coli* and used to again co-transfect COS cells. This can be repeated three times, if necessary, thus repeatedly enriching for positive clones. Individual colonies can be tested by co-transfection with pIL-8 and pCMV and confocal microscopy. Double positive clones are eliminated and the pIL8 (+)/pCMV(-) clones are chosen for further study.

Example 5. Design of Pool Screens

In the event that the above techniques are not readily available, one may still employ pool screening techniques with the expression cloning method described herein. The data presented above show that it will be possible to identify from a pool of cDNAs, a single cDNA encoding a cellular component controlling the expression of GFP via the pIL-8 construct. In previous applications of mammalian expression cloning techniques, the search has been for a unique protein, for example a receptor subunit. However in the experiments proposed here multiple gene products can be expected to read out in the screen.

This fact raises two issues that need to be addressed:

i) Elimination of false positives:

To eliminate from further analysis those cDNAs that affect steps other than the rate of transcription and lead to GFP accumulation. These include plasmid replication, mRNA stability, rate of translation, and rate of GFP degradation. While many of the cDNAs that read out for these reasons may be of interest, within the scope of the proposed studies they will be taken as "false positives." These can be screened out by testing against the pCMV reporter and eliminated if

they enhance expression, because even though the CMV promoter contains several mammalian transcription factor binding sites in the 600 bp core fragment, it is constitutively active in COS-7 cells and is unaffected by IL-1, for example.

ii) Optimization of Pool Size

It is important to optimize the pool versus library size in order to minimize the time taken to screen the entire library while maximizing the discriminatory power of the screen, bearing in mind that signal strength, as measured by positive cells per pool transfected, increases linearly with decreasing pool size.

Some estimates of the overall "behavior" of the library may be made from general information about eukaryotic message levels and also from our previous experience. General estimates suggest that translationally active and proliferating eukaryotic cells express about 400,000 copies per cell of polyA+ MRNA. A "rare" mRNA is expressed at about 10-20 copies per cell, and therefore about one in 30,000 mRNAs would be found to encode any particular rare protein (~100,000) protein molecules / cell).

We can compare these estimates with the results from our cloning of IL-1-RII (McMahon, C., *et al.* (1991) *EMBO J.* 10:2821-2832). A library of 5 X 10^6 cDNAs was prepared from polyA+RNA (from CB23 cells: expressing about 20,000 copies per cell of the protein) using random hexanucleotide primers and screened in pools of 4,000 cDNAs. One of the first 85 pools tested by CV-1-EBNA transfection and ^{125}IL-1 binding autoradiography was positive and the screen was then terminated. Thus, one functional transcript was detected in 9% of the library, suggesting that there were about 11 such transcripts in the entire library, or one in 350,000.

This would be consistent with the expected mRNA frequency if only 10% of the relevant message was converted to cDNAs that gave functional transcripts.

Since the coding region and the 3' UTR of the IL-1RI mRNA are of approximately equal length (about 1.2 Kb) this suggests that half the transcripts would have primed 5' to the stop codon and hence failed to give functional mRNAs because our vector lacked a universal terminator. Bearing in mind the semi quantitative nature of the above estimates, this leaves a factor of 5 which could be accounted for by premature termination during cDNA synthesis and destabilizing sequences in the 3' UTR.

These calculations suggest that, in a library made from polyA+RNA using random hexanucleotide primers, an estimated one in 300,000 cDNAs will express a functional protein product corresponding to a unique gene expressed at low levels. As an illustration, the results of a simulated screen of 10^6 cDNAs, in pools of either 10,000 or 1,000, assuming 5 functional cDNAs per unique mRNA, as the number of unique cDNAs that read out in the screen rises from 1 to 10,000 is shown in Table I. The figure of 10 is likely to be a realistic minimum in our screens because there are already known to be at least this number of components controlling the three known transcription factors with binding sites in pIL-8. The true number of positives cannot be reasonably estimated, since i) we do not know how many unknown components of the relevant signal transduction pathways exist, and ii) while it is likely that "signal transduction systems are degenerate," we do not know how degenerate.

Since the initial screen is likely a minor component of the work in the project, it would be prudent to test a small number of pools (20) at both pool sizes against pIL-8±-IL-1 and pCMV±IL-1 (see above) to determine both the number of positive pools and the number of positive cells per pool versus pool size. These data will allow us to estimate how many cDNA clones are likely to be isolated and to choose an optimal pool size.

100 pools of 10,000								1000 pools of 1,000							
cDNAs per Library	1	3	10	30	100	300	1000	cDNAs per Library	1	3	10	30	100	300	1000
cDNAs per pool								cDNAs per pool							
0	95	86	61	22	1	0	0	0	995	985	951	861	607	223	7
1	5	13	30	34	3	0	0	1	5	15	48	129	303	335	33
2	0	1	8	25	8	0	0	2	0	0	1	10	76	252	84
3	0	0	1	13	14	0	0	3	0	0	0	0	13	126	140
4	0	0	0	4	18	0	0	4	0	0	0	0	2	47	176
5	0	0	0	2	18	0	0	5	0	0	0	0	0	14	176
6	0	0	0	0	15	0	0	6	0	0	0	0	0	4	147
7	0	0	0	0	11	1	0	7	0	0	0	0	0	1	105
8	0	0	0	0	6	2	0	8	0	0	0	0	0	0	66
9	0	0	0	0	4	3	0	9	0	0	0	0	0	0	36
10	0	0	0	0	2	4	0	10	0	0	0	0	0	0	18
10	0	0	0	0	0	90	100	10	0	0	0	0	0	0	13

Table I: Number of Positive Pools as a Function of Number of Positive cDNAs in Library. Entries show the predicted number of pools expressing a given number of positive cDNAs, calculated using the binomial theorem. The approximate number of positive cells/10^5 can be estimated by multiplying the cDNAs per pool column by 30 for the 10,000 pool size and 300 for the 1,000 pool size. The cDNAs per library row represents unique cDNAs, each present at 5 copies in the library.

Example 6. Strategic Considerations for Future Experiments

Additional experiments can be performed using functional cloning and single cell detection system of the invention. These experiments include:

(1) Isolation of novel cDNA clones which encode elements of the cytokine signal transduction systems. cDNA clones that modulate the test IL-8 promoter can be isolated by the direct expression cloning methods described above. The cDNA library is prepared from CV-1 cells. This will help to avoid any species incompatibility problems between over expressed proteins and the background signaling machinery in COS cells.

(2) Mapping the functions of the products of all cDNAs isolated against the test promoter using a series of test promoter deletion constructs, containing all six pairwise and single binding site combinations. For example, the following promoter combinations can be tested and should define the sites through which, for any cloned sequence, IL-8 activity is modulated.

	AP1	c/EBP	NF-κβ
wild type	+	+	+
Δ NF-κβ	+	+	-
Δ EBP	+	-	+
Δ AP1	-	+	+
ΔΔ NF-κβ/EBP	+	-	-
ΔΔ NF-κβ/AP1	-	+	-
ΔΔ AP1/EBP	-	-	+
ΔΔΔ NF-κβ/EBP/AP1	-	-	-

(3) Elucidating pathways by stimulating cells transfected with pIL-8± novel cDNAs with IL-1 and both, to map which elements that lie on pathways from TNFR, IL-1R and which are common.

(4) Further elucidating pathways by combinatorial transfections with cloned cDNAs and stimulating with TNF, IL-1, and both, to determine quantitatively which cDNA regulates the most rate critical steps in the pathways and in which order the steps are organized by measuring the rates of GFP accumulation.

The methods we will use have been described in the previous sections. We are, as the pilot data illustrate, already performing GFP transfections and measurements on a routine basis. These experiments will allow us to clone cDNAs encoding novel signal transduction pathway elements, to determine which are most critical for setting the rate of pathway activity, to determine which pathways they control, and to gain some initial ideas of to what extent the pathways for IL-1 and TNF overlap. This information can all be generated within the experimental framework outlined above. In many instances the sequence of each cDNA clone will provide a clear cut guide as to its function, based on homology to known families of molecules. In addition, the frequency of positive clones in our libraries will allow us to estimate the number of different cellular components which can impact on the activity of this limited set of pathways. This will allow us to begin to address broader issues of redundancy and cross talk in the network of interacting signals that control patterns of gene expression in mammalian cells.

Although we have demonstrated our functional cloning and single cell detection methodology with respect to the cytokine signal transduction pathway, in theory the system could be used to isolate any sequence that modulates the

transcription of a test promoter or other regulatory gene sequence. Therefore, it would be possible to isolate those sequences important in the control of the P selectin, ELAM-1, and CD 11a promoters, to name a few.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

All articles and patents cited herein are incorporated by reference and repeated here for convenience only:

References

1. Fanslow, W., *et al.* 1990 *Science* **248**:739-742.

2. MoWer, K., D. *et al.* 1993 *J. Immunol.* **151**:1548-1561.

3. Sims, J., *et al.* 1988 *Science* **241**:585-589.

4. Smith, C., D. *et al.* 1990 *Science* **248**:1019-1023.

5. Gruss, H. J., and Dower, S.K. 1995 *Cytokines and Molecular Therapy* **1**:75-105.

6. Sims, J. E., and Dower, S.K. 1994 *European Cytokine Network 5:539-546.*

7. Rothe, M. , 1994 *Cell* **78**:681-692.

8. Karin, M., and Smeal, T. 1992 *Trends In Biochem. Sci.* **17**:418-422.

9. Sha. W.H. *el al.* 1995 *Cell* **80**:321-330.

10. Huber, A., *et al.* 1991 *Science* **254**:99-102.

11. Sims, J., et al. 1994 IL-1 SIGNAL TRANSDUCTION *in Signal Transduction Through Growth Factor Receptors., vol. 3.* JAI Press, Greenwich, Conn., USA, London, U,K., p. 197-222.

12. Mitcham, J., *et al.* 1996 *J Biol. Chem.* **271**:5777-5783.

13. Dower, S. 1993 P PHARMACOLOGICAL, BIOCHEMICAL AND MOLECULAR BIOLOGICAL STUDIES ON CYTOKINE RECEPTORS *in Neurobiology of Cytokines, vol. 16.*, E. B. De Souza, ed. Academic Press, Inc., New York, p. 3-30.

14. Fields, S. & Song, 0. 1989 *Nature* **340**:245-6.

15. Seed, B. 1995 *Current Opinion In Biotechnology* **6**:567-573.

16. Davis, S., *et al.* 1996 *Cell* **87**:1161-1169.

17. Chalfie, M., *et al.* 1994 *Science* **263**:802-805.

18. McMahon, C., *et al.* 1991 *EMBO J.* **10**:2821-2832.

19. Valdivia, R.H., Falkow, S., 1996 *Mol. Microbiol.* **22(2)**:367-78.

20. Holding, C. & Monk, M. 1990 Genet. Res. **55(2)**:120.

21. Yamasaki, *et al.* (1988) *Science,* **241**:825-28

CLAIMS

We claim:

1. A method of determining whether a the expression from a test promoter or other gene regulatory sequence is modulated by a test nucleic acid or test compound , comprising:

a. transfecting a host cell with a first nucleic acid comprising a first test promoter and/or other gene regulatory sequence coupled with a nucleotide sequence encoding a detectable reporter protein, to produce a transfected cell;

b. providing to the transfected cell (i) a second nucleic acid comprising a second promoter coupled to a test nucleotide sequence, to produce a co-transfected cell or (ii) a test compound; and

c. detecting the expression of said reporter protein in said co-transfected cell if the expression levels of said reporter protein is different from the level found in a cell transfected only with said first nucleic acid sequence.

2. The method of claim 1, wherein said host cell is a mammalian cell.

3. The method of claim 2, wherein said mammalian cell is a COS or a Hela cell.

4. The method of claim 1, wherein said reporter protein is fluorescently, radioactively, biologically and/or enzymatically labeled.

5. The method of claim 1, wherein said reporter protein is selected from the group consisting of green fluorescent protein (GFP), EGFP™, d2-EGFP, S6ST, T203Y, BFP and GFPmut or any other GFP variant.

6. The method of claim 1, wherein said reporter gene expression is detected by fluorescence activated cell sorting or fluorescence microscopy.

7. The method of claim 1, wherein said test promoter comprises at least a portion of a cytokine promoter.

8. The method of claim 7, wherein said test promoter comprises at least a portion of a promoter selected from the group consisting of an IL-8, a P-selectin, an ELAM-1 and a CD11 promoter.

9. The method of claim 7, wherein said cytokine promoter is modulated by IL-1 or TNFα other components that directly or indirectly induce the IL-8 promoter.

10. The method of claim 1, wherein said test nucleotide sequence modulates one or more of the rate of transcription, plasmid replication, mRNA stability, rate of translation and rate of degradation of said detectable reporter protein.

11. A method for determining whether the expression of a test promoter or other gene regulatory sequence is modulated by a test nucleic acid in response to one or more agents, comprising:

a. transfecting a host cell with a first nucleic acid comprising a first test promoter and/or other gene regulatory sequence coupled with a nucleotide sequence encoding a detectable reporter protein, to produce a transfected cell;

b. transfecting said transfected cell with a second nucleic acid comprising a second promoter coupled to a test nucleotide sequence, to produce a co-transfected cell;

c. providing to said co-transfected cell one or more agents; and

d. detecting the expression of said reporter protein in said agent-stimulated co-transfected cell if the expression levels of said reporter protein is different from the level found in said co-transfected cell that was not provided with said agent(s).

12. The method of claim 11, wherein said host cell is a mammalian cell.

13. The method of claim 12, wherein said mammalian cell is a COS or a Hela cell.

14. The method of claim 11, wherein said reporter protein is fluorescently, radioactively, biologically and/or enzymatically labeled.

15. The method of claim 14, wherein said reporter protein is selected from the group consisting of green fluorescent protein (GFP), EGFP™, d2-EGFPS6ST, T203Y, BFP and GFPmut or other GFP mutant.

16. The method of claim 14, wherein said reporter gene expression is detected by fluorescence activated cell sorting or fluorescence microscopy.

17. The method of claim 11, wherein said test promoter comprises at least a portion of a cytokine promoter.

18. The method of claim 11, wherein said test promoter comprises at least a portion of a promoter selected from the group consisting of an IL-8, a P selectin, an ELAM-1 and a CD11promoter.

19. The method of claim 11, wherein said agent(s) participate in a cytokine signal transduction pathway.

20. The method of claim 11, wherein said agent is a cytokine selected from the group consisting of IL-1 and TNFα.

21. The method of claim 11, wherein said expression levels are increased in response to said agent(s).

22. The method of claim 11, wherein said expression levels are decreased in response to said agent(s).

23. The method of claim 11, further comprising the step of selecting against molecules that stimulate expression of said reporter protein when coupled to a test promoter which is unresponsive to said agent(s), thereby eliminating false positives if expression is enhanced.

24. A method of identifying and isolating from a live cell a nucleic acid encoding a molecule that modulates the expression of a test promoter or other gene regulatory sequence, comprising the steps of:

a. transfecting a host cell with a nucleic acid comprising a test promoter and/or other gene regulatory sequence coupled with a nucleotide sequence encoding a detectable reporter protein, to produce transfected cells;

b. transfecting said transfected cells with a library of cDNA molecules, to produce a co-transfection library;

c. detecting the expression of said reporter protein in said co-transfection library;

d. sorting the co-transfection library to create an enriched co-transfection library, wherein said co-transfected cells are selected for inclusion of cells expressing said reporter protein at levels that are different from those found in unsorted co-transfected cells;

e. purifying, transforming, amplifying and repurifying the cDNA molecules from step (b) until a sufficient level of enrichment for cDNA molecules that modulate the expression of said reporter is obtained; and

f. isolating said cDNA molecules.

25. The method of claim 24, wherein said host cell is a mammalian host cell.

26. The method of claim 25, wherein said mammalian cell is a COS or a Hela cell.

27. The method of claim 24, wherein said reporter protein is fluorescently, radioactively, biologically and/or enzymatically labeled.

28. The method of claim 27, wherein said reporter protein is selected from the group consisting of green fluorescent protein (GFP), EGFP™, d2-EGFP, S6ST, T203Y and GFPmut or other GFP variant.

29. The method of claim 27, wherein said reporter gene expression is detected by fluorescence activated cell sorting or fluorescence microscopy.

30. The method of claim 24, wherein said test promoter comprises at least a portion of a cytokine promoter.

31. The method of claim 24, wherein said test promoter comprises at least a portion of a promoter selected from the group consisting of an IL-8, a P selectin, an ELAM-1 and a CD11promoter.

32. A method of identifying and isolating from a live cell a nucleic acid encoding a molecule that modulates expression from a test promoter or other gene regulatory sequence in response to one or more agents, comprising the steps of :

a. transfecting a host cell with a nucleic acid comprising a test promoter and/or other gene regulatory sequence coupled with a nucleotide sequence encoding a detectable reporter protein, to produce transfected cells;

b. transfecting said transfected cells with a library of cDNA molecules, to produce a co-transfection library;

c. providing to said co-transfection library one or more agents;

d. detecting the expression of said reporter protein in said co-transfection library;

e. sorting the co-transfection library to create an enriched co-transfection library, wherein said co-transfected cells are selected for inclusion of cells expressing said reporter protein at levels that are different from those found in unsorted co-transfected cells;

f. purifying, transforming, amplifying and repurifying the cDNA molecules from step (b) until a sufficient level of enrichment for cDNA molecules that modulate the expression of said reporter is obtained; and

g. isolating said cDNA molecules.

33. The method of claim 32, wherein said host cell is a mammalian cell.

34. The method of claim 33, wherein said mammalian cell is a COS or a Hela cell.

35. The method of claim 32, wherein said reporter protein is fluorescently, radioactively, biologically and/or enzymatically labeled.

36. The method of claim 35, wherein said reporter protein is selected from the group consisting of green fluorescent protein (GFP), EGFP™, d2-EGFP, S6ST, T203Y, GFPmut or other GFP variant.

37. The method of claim 35, wherein said reporter gene expression is detected by fluorescence activated cell sorting or fluorescence microscopy.

38. The method of claim 32, wherein said test promoter comprises at least a portion of a cytokine promoter.

39. The method of claim 38, wherein said test promoter comprises at least a portion of a promoter selected from the group consisting of an IL-8, a P selectin, an ELAM-1 and a CD11promoter.

40. The method of claim 32, wherein said agent(s) participate in a cytokine signal transduction pathway.

41. The method of claim 32, wherein said agent is a cytokine selected from the group consisting of IL-1 or TNFα.

42. The method of claim 32, wherein said expression levels are increased in response to said agent(s).

43. The method of claim 32, wherein said expression levels are decreased in response to said agent(s).

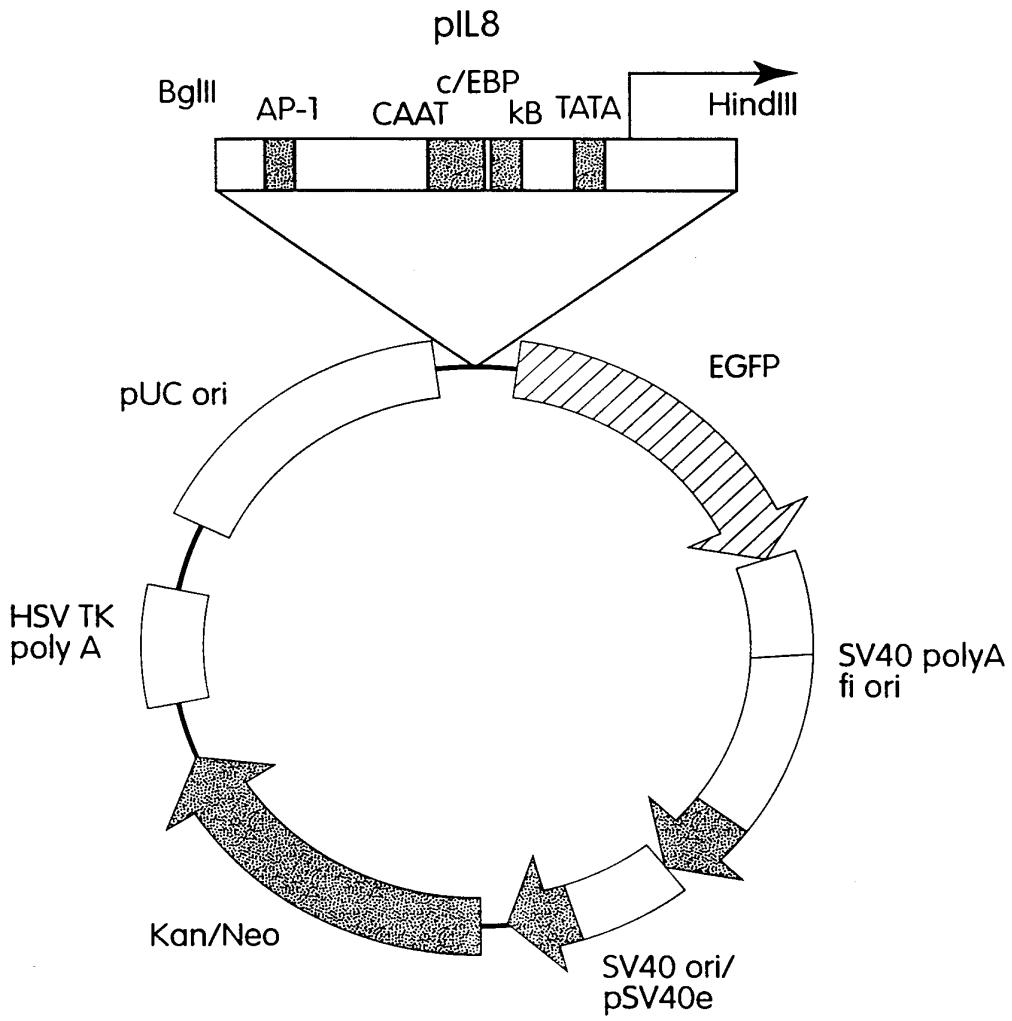

pIL8
BglII
AP-1
c/EBP
CAAT
kB
TATA
HindIII
pUC ori
EGFP
HSV TK
poly A
SV40 polyA
fi ori
Kan/Neo
SV40 ori/
pSV40e

Fig. 1

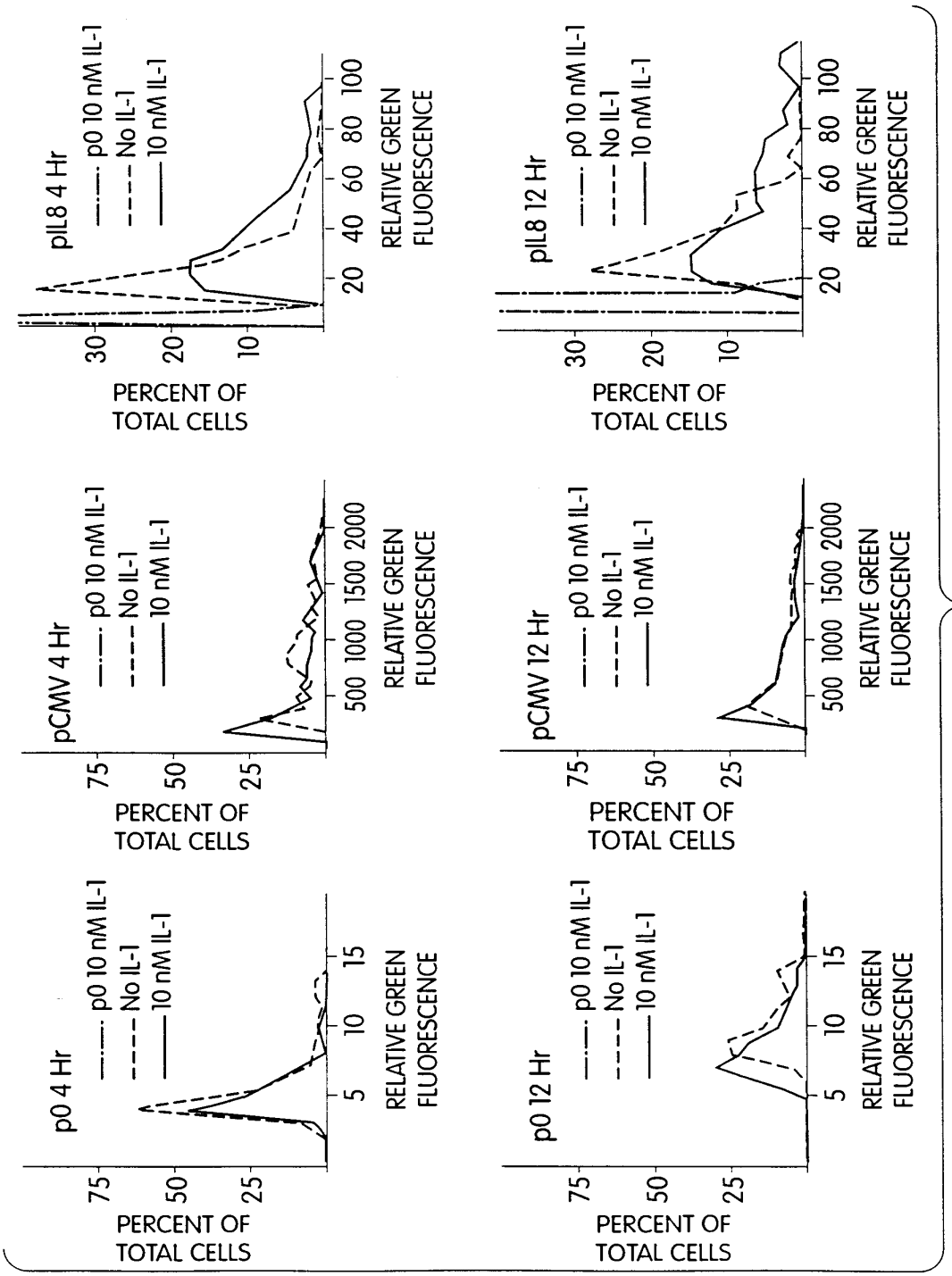

pIL8 4 Hr
p0 10 nM IL-1
No IL-1
10 nM IL-1
PERCENT OF TOTAL CELLS
RELATIVE GREEN FLUORESCENCE
pIL8 12 Hr
pCMV 4 Hr
pCMV 12 Hr
p0 4 Hr
p0 12 Hr

Fig. 2

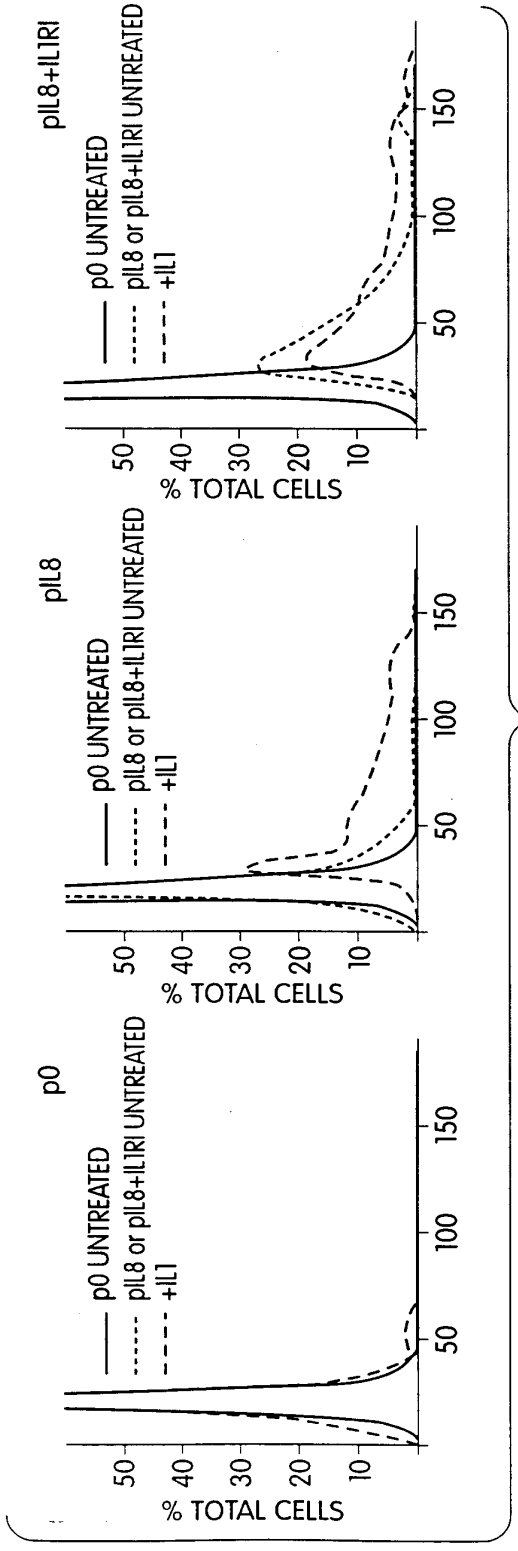

pIL8+IL1RI
p0 UNTREATED
pIL8 or pIL8+IL1RI UNTREATED
+IL1
% TOTAL CELLS
pIL8
p0 UNTREATED
pIL8 or pIL8+IL1RI UNTREATED
+IL1
% TOTAL CELLS
p0
p0 UNTREATED
pIL8 or pIL8+IL1RI UNTREATED
+IL1
% TOTAL CELLS

Fig. 3

INTERNATIONAL SEARCH REPORT

Int. tional Application No
PCT/US 98/26715

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 88 03168 A (SALK INST FOR BIOLOGICAL STUDI) 5 May 1988 see whole doc. esp. claims 28 ff. ---	1-3, 11-13
X	WO 96 01899 A (SCHERING CORP) 25 January 1996 see whole docs. esp. claims ---	1-3, 11-13
A	US 5 569 588 A (ASHBY MATTHEW ET AL) 29 October 1996 see whole doc., esp. claims --- -/--	1-43

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 April 1999	27/04/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Müller, F

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. .tional Application No
PCT/US 98/26715

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KAIN S R ET AL: "GREEN FLUORESCENT PROTEIN AS A REPORTER OF GENE EXPRESSION AND PROTEIN LOCALIZATION" BIOTECHNIQUES, vol. 19, no. 4, October 1995, pages 650-655, XP002033687 see the whole document ---	1-43
A	MITCHAM J.L. ET AL.,: "T1/ST2 signaling establishes it as a member of an expanding interleukin-1 receptor family" J. BIOL. CHEMISTRY, vol. 271, no. 10, - 8 March 1996 pages 5777-5783, XP002099578 cited in the application see esp. p.5778, 2.col., 2.par.; p.5781,1.col., 1.par. ---	
P,X	WO 98 39483 A (VENTANA GENETICS INC) 11 September 1998 see whole doc. esp. claims and examples, p.23 ff -----	1-43

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. ational Application No
PCT/US 98/26715

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 8803168 A	05-05-1988	US 5071773 A	10-12-1991
		AU 616389 B	31-10-1991
		AU 8238887 A	25-05-1988
		EP 0287653 A	26-10-1988
		EP 0733705 A	25-09-1996
		JP 10004987 A	13-01-1998
		JP 10108684 A	28-04-1998
		JP 1500964 T	06-04-1989
		US 5597705 A	28-01-1997
		US 5534418 A	09-07-1996
		US 5606021 A	25-02-1997
		US 5312732 A	17-05-1994
		US 5298429 A	29-03-1991
WO 9601899 A	25-01-1996	AU 2946695 A	09-02-1996
		CA 2194361 A	25-01-1996
		EP 0770130 A	02-05-1997
		JP 10502535 T	10-03-1998
US 5569588 A	29-10-1996	AU 6720996 A	05-03-1997
		CA 2202154 A	20-02-1997
		EP 0791078 A	27-08-1997
		JP 10507647 T	28-07-1998
		WO 9706277 A	20-02-1997
WO 9839483 A	11-09-1998	AU 6543898 A	22-09-1998

Form PCT/ISA/210 (patent family annex) (July 1992)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 January 2003 (16.01.2003)

PCT

(10) International Publication Number
WO 03/003810 A2

(51) International Patent Classification: Not classified

(21) International Application Number: PCT/US02/21278

(22) International Filing Date: 3 July 2002 (03.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/898,743 3 July 2001 (03.07.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/898,743 (CIP)
Filed on 3 July 2001 (03.07.2001)

(71) Applicant *(for all designated States except US)*: **THE INSTITUTE FOR SYSTEMS BIOLOGY** [US/US]; 1441 North 34th Street, Seattle, WA 98103-8904 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **DIMITROV, Krassen** [BG/US]; 10759 Burke Avenue N., Seattle, WA 98133 (US). **DUNAWAY, Dwayne** [US/US]; 6548 39th Avenue N.E., Seattle, WA 98115 (US).

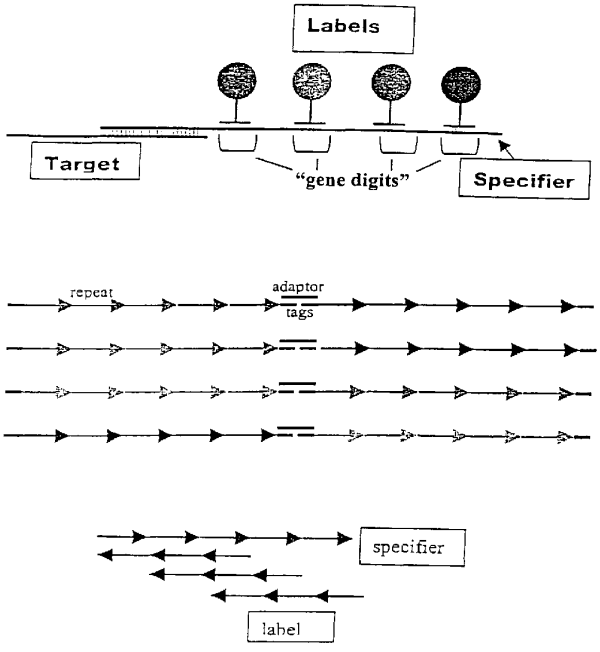
(74) Agents: **CADENA, Deborah, L.** et al.; Campbell & Flores LLP, 7th floor, 4370 La Jolla Village Drive, San Diego, CA 92122 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,

[Continued on next page]

(54) Title: METHODS FOR DETECTION AND QUANTIFICATION OF ANALYTES IN COMPLEX MIXTURES



(57) Abstract: The invention provides a diverse population of uniquely labeled probes, containing about thirty or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid. Also provided is a method of producing a population of uniquely labeled nucleic acid probes. The method consists of (a) synthesizing a population of target specific nucleic acid probes each having a different specifier; (b) synthesizing a corresponding population of anti-genedigits each having a unique label, the population having a diversity sufficient to uniquely hybridize to genedigits within the specifiers, and (c) hybridizing the populations of target nucleic acid probes to the anti-genedigits, to produce a population in which each of the target specific probes is uniquely labeled. Also provided is a method of detecting a nucleic acid analyte. The method consists of (a) contacting a mixture of nucleic acid analytes under conditions sufficient for hybridization with a plurality of target specific nucleic acid probes each having a different specifier; (b) contacting the mixture under conditions sufficient for hybridization with a corresponding plurality of anti-genedigits each having a unique label, the plurality of anti-genedigits having a diversity sufficient to uniquely hybridize to genedigits within the specifiers, and (c) uniquely detecting a hybridized complex between one or more analytes in the mixture, a target specific probe, and an anti-genedigit.

WO 03/003810 A2

TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS FOR DETECTION AND QUANTIFICATION OF ANALYTES IN COMPLEX MIXTURES

BACKGROUND OF THE INVENTION

This invention relates generally to the field of genomics and, more specifically to detection, identification, and quantification of target analytes in mixtures.

Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression are at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in normal cells compared to diseased cells can aid the discovery of new drug and diagnostic targets.

Nucleic acids can be detected and quantified based on their specific polynucleotide sequences. The basic principle underlying existing methods of detection and quantification is the hybridization of a labeled complementary probe sequence to a target sequence of interest in a sample. The formation of a duplex indicates the presence of the target sequence in the sample and the degree of duplex formation, as measured by the amount of label incorporated in it, is proportional to the amount of the target sequence.

This technique, called molecular hybridization, has been a useful tool for identifying and analyzing specific nucleic acid sequences in complex mixtures. This technique has been used in diagnostics, for example, to detect nucleic acid sequences of various microbes in

biological samples. In addition, hybridization techniques have been used to map genetic differences or polymorphisms between individuals. Furthermore, these techniques have been used to monitor changes in gene expression in different populations of cells or in cells treated with different agents.

In the past, only a few genes could be detected in a complex sample at one time. However, DNA microarrays, devices that consist of thousands of immobilized DNA sequences present on a miniaturized surface, have made this process more efficient. Using a microarray, it is possible in a single experiment to detect the presence or absence of thousands of genes in a biological sample. This allows researchers to simultaneously perform several diagnostic tests on one sample, or to observe expression level changes in thousands of genes in one experiment. Generally, microarrays are prepared by binding DNA sequences to a surface such as a nylon membrane or glass slide at precisely defined locations on a grid. Then nucleic acids in a biological sample are labeled and hybridized to the array. The labeled sample DNA marks the exact position on the array where hybridization occurs, allowing automatic detection.

Unfortunately, despite the miniaturization of array formats, this method still requires significant amounts of the biological sample. However, in several cases, such as biopsies of diseased tissues or samples of a discrete cell type, the biological sample is in limited supply. In addition, the kinetics of hybridization on the surface of a microarray is less efficient than hybridization in small amounts of aqueous solution. Furthermore, microarrays require a large dynamic range of detection to account for large difference in abundance of the different molecular species. This results in

decreased sensitivity since there is a trade-off between sensitivity and dynamic range. A further problem with microarray methods is that the output is quantitative analog data that has undergone several intermediary transformations. In microarrays, the amount of nucleic acid hybridized to each spot is determined by measuring its label and so any nonlinear correlation between the amount of DNA hybridized and the amount of the label detected will skew the data output. Such non-linearity has been widely documented.

Thus, there exists a need for accurate and sensitive detection, identification and quantification of analytes in complex mixtures. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

The invention provides a diverse population of uniquely labeled probes, containing about thirty or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid. Also provided is a method of producing a population of uniquely labeled nucleic acid probes. The method consists of (a) synthesizing a population of target specific nucleic acid probes each having a different specifier; (b) synthesizing a corresponding population of anti-genedigits each having a unique label, the population having a diversity sufficient to uniquely hybridize to genedigits within the specifiers, and (c) hybridizing the populations of target nucleic acid probes to the anti-genedigits, to produce a population in which each of the target specific probes is uniquely labeled. Also provided is a method of detecting a nucleic acid analyte. The method consists of (a) contacting a mixture of nucleic acid analytes under conditions sufficient for

hybridization with a plurality of target specific nucleic acid probes each having a different specifier; (b) contacting the mixture under conditions sufficient for hybridization with a corresponding plurality of anti-genedigits each having a unique label, the plurality of anti-genedigits having a diversity sufficient to uniquely hybridize to genedigits within the specifiers, and (c) uniquely detecting a hybridized complex between one or more analytes in the mixture, a target specific probe, and an anti-genedigit.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows components of a specifier. Figure 1A shows an association between a target and a labeled specifier. Figure 1B shows a structure of a genedigit. Figure 1C shows an association between a genedigit and labeled anti-genedigit.

Figure 2 shows a DNA probe molecule (specifier) that contains several different types of spatially separated labels.

Figure 3 shows DNA probe molecules containing a label on one end that have been stretched on a coverslip using a flow-stretch technique.

Figure 4 shows labeled DNA molecules that have been aligned using an electro-stretch technique. Figure 4A shows labeled DNA probe molecules in a flow cell with the voltage off. Figure 4B shows the labeled DNA probe molecule in the flow cell with the voltage on.

Figure 5 shows a schematic of probes being aligned by constriction in the flow of a liquid containing the probes in the presence of an oscillating electric field.

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to methods for the generation of a diverse population of unique labels that can be used for the detection, identification, and direct quantification of a wide variety of target analytes. The methods are advantageous in that they generate a large number of unique labels of about the same unit signal starting from just a small number of different labels. Enough labels are generated by this method so that each analyte in a complex mixture can be uniquely bound by a label and thus identified. The labels are designed so that they can be used in a small volume of solution which increases the efficiency of the binding reaction and is useful when only small amounts of sample are available for analysis. After the individual molecules in a sample are identified, they can be directly counted resulting in a digital read-out of each molecular species in a mixture. Since the diverse labels are based on a small number of starting labels, this allows the detection method to operate in a narrow dynamic range resulting in improved sensitivity of the system since the trade-off between sensitivity and dynamic range is avoided. The methods of the invention therefore provide an accurate and sensitive system for the detection, identification and quantitation of analytes in a mixture.

In one embodiment, the invention is directed to detecting nucleic acid analytes in a complex mixture by first contacting the mixture under conditions sufficient for hybridization with a plurality of target specific nucleic acid probes. These target specific nucleic acid probes, called specifiers, contain a target specific region and a region containing one or more unique "genedigit" sequences. The genedigits consist of repeated core element sequences that can be specifically

bound by a complementary anti-genedigit sequence which can contain a unique label. The mixture containing the nucleic acid analytes and the specifiers is then contacted with a corresponding plurality of labeled anti-genedigits having a diversity sufficient to uniquely hybridize to genedigits within the specifiers. This allows the unique detection of a hybridized complex between analytes in the mixture and specifiers with unique labels.

As used herein, the term "bound" when referring to a unique label or nucleic acid is intended to mean that a label monomer is attached to a nucleotide in a 1:1 correspondence. A label monomer as used herein is intended to mean an individual measurable moiety, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other moiety known in the art that is measurable by analytical methods. A label monomer can be directly attached to a nucleotide using methods well known in the art. Nucleotides can also be chemically modified or derivitized in order to attach a label monomer. For example, a fluorescent monomer such as a fluorescein molecule can be attached to dUTP (deoxyuridine-triphosphate) using a four-atom aminoalkynyl group. Each label monomer is attached to a nucleotide making a label monomer:nucleotide complex. This label monomer:nucleotide can be incorporated into nucleic acids in a variety of ways. For example, a label monomer:nucleotide can be incorporated at only one location within a nucleic acid or at two or more locations within a nucleic acid. A nucleotide can be attached to a label monomer first and then be incorporated into a nucleic acid, or an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid. In addition, for example, a label monomer:nucleotide can be incorporated

into a nucleic acid and a different type of label monomer:nucleotide can be incorporated into the same nucleic acid.

As used herein, "analyte" or target when referring to an analyte is intended to mean any molecule whose presence is measured. An analyte molecule can be essentially any molecule for which a detectable probe or assay exists or can be produced by one skilled in the art. For example, an analyte can be a macromolecule such as a nucleic acid, polypeptide or carbohydrate, or a small organic compound. Measurement can be quantitative or qualitative. An analyte can be part of a sample that contains other components or can be the sole or major component of the sample. Therefore, an analyte can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The analyte can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also the analyte can have either a known or unknown structure or sequence.

As used herein, the term "target specific" is intended to mean an agent that binds to a target analyte selectively. This agent will bind with preferential affinity toward the target while showing little to no detectable cross-reactivity toward other molecules. For example, when the target is a nucleic acid, a target specific sequence is one that is complementary to the sequence of the target and able to hybridize to the target sequence with little to no detectable cross-reactivity with other nucleic acid molecules. A nucleic acid target could also be bound in a target specific manner by a protein, for example by the DNA binding domain of a transcription factor. If the target is a

protein or peptide it can be bound specifically by a nucleic acid aptamer, or another protein or peptide, or by an antibody or antibody fragment which are sub-classes of proteins.

The term "complementary" refers to two nucleotides that can form multiple thermodynamically favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if the nucleotides of the first sequence are complementary to the nucleotides of the second sequence. The percent of complementarity (i.e. how many nucleotides from one strand form multiple thermodynamically favorable interactions with the other strand compared with the total number of nucleotides present in the sequence) indicates the extent of complementarity of two sequences.

As used herein, the term "repeat sequence" is intended to mean two or more copies of a core element. A repeat sequence can have direct repetition of the core sequence without any intervening sequence, or the repeat sequence can have non-consecutive repetition of the core element with intervening sequences. A core element can be made of nucleic acids such as an oligonucleotide or an aptamer, or the core element can be made of amino acids such as a peptide sequence. If, for example, the core element is a nucleic acid sequence of 8 base pairs then three direct repeats of this sequence would be a 24 base sequence. A "complimentary repeat sequence" is a sequence that binds specifically to the repeat sequence. For the example above where the repeat sequence is repetition of a nucleic acid core element, the complimentary repeat sequence can contain one or more

copies of the complementary strand of the core element that will hybridize specifically to the repeat sequence.

As used herein, the term "genedigit" is intended to mean a region of pre-determined nucleotide or amino acid sequence that serves as an attachment point for a label. The genedigit can have any structure including, for example, a single unique sequence or a sequence containing repeated core elements. Each genedigit has a unique sequence which differentiates it from other genedigits. An "anti-genedigit" is a nucleotide or amino acid sequence or structure that binds specifically to the gene digit. For example, if the genedigit is a nucleic acid, the anti-genedigit can be a nucleic acid sequence that is complementary to the genedigit sequence. If the genedigit is a nucleic acid that contains repeated core elements then the anti-genedigit can be a series of repeat sequences that are complementary to the repeat sequences in the genedigit. An anti-genedigit can contain the same number, or a lesser number, of repeat sequences compared to the genedigit as long as the anti-genedigit is able to specifically bind to the genedigit.

As used herein, the term "specifier" is intended to mean the linkage of one or more genedigits to a target specific sequence. The genedigits can be directly linked or can be attached using an intervening or adapting sequence. A specifier can contain a target specific sequence which will allow it to bind to a target analyate. An "anti-specifier" has a complementary sequence to all or part of the specifier such that it specifically binds to the specifier.

As used herein, the term "mixture" is intended to mean a composition that contains more than one molecule. A mixture can be homogenous, containing a single species,

or heterogeneous, containing different species. Examples of homogeneous samples include, for example, isolated populations of polypeptides, nucleic acids or carbohydrates. Heterogeneous mixtures include extracts from tissues, cells, lysates and fractionated portions thereof. For example, a mixture can be a pure solution containing several molecules of a single protein, or a mixture can be an extract from a cell containing several proteins and other types of macromolecules.

As used herein, the term "label" is intended to mean a molecule or molecules that render an analyte detectable by an analytical method. An appropriate label depends on the particular assay format and are well known by those skilled in the art. For example, a label specific for a nucleic acid molecule can be a complementary nucleic acid molecule attached to a label monomer or measurable moiety, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other moiety known in the art that is measurable by analytical methods. In addition, a label can include any combination of label monomers.

As used herein, "unique" when used in reference to label is intended to mean a label that has a detectable signal that distinguishes it from other labels in the same mixture. Therefore, a unique label is a relative term since it is dependent upon the other labels that are present in the mixture and the sensitivity of the detection equipment that is used. In the case of a fluorescent label, a unique label is a label that has spectral properties that significantly differentiate it from other fluorescent labels in the same mixture. For example, a fluorescein label can be a unique label if it is included in a mixture that contains a rhodamine label since these fluorescent labels emit light at distinct,

essentially non-overlapping wavelengths. However, if another fluorescent label was added to the mixture that emitted light at the same or very similar wavelength to fluorescein, for example the Oregon Green fluorophore, then the fluorescein would no longer be a unique label since Oregon Green and fluorescein could not be distinguished from each other. A unique label is also relative to the sensitivity of the detection equipment used. For example, a FACS machine can be used to detect the emission peaks from different fluorophore-containing labels. If a particular set of labels have emission peaks that are separated by, for example, 2 nm these labels would not be unique if detected on a FACS machine that can distinguish peaks that are separated by 10nm or greater, but these labels would be unique if detected on a FACS machine that can distinguish peaks separated by 1nm or greater.

As used herein, the term "signal" is intended to mean a detectable, physical quantity or impulse by which information on the presence of an analyte can be determined. Therefore, a signal is the read-out or measurable component of detection. A signal includes, for example, fluorescence, luminescence, colorimetric, density, image, sound, voltage, current, magnetic field and mass. Therefore, the term "unit signal," as used herein is intended to mean a specified quantity of a signal in terms of which the magnitudes of other quantities of signals of the same kind can be stated. Detection equipment can count signals of the same type and display the amount of signal in terms of a common unit. For example, a nucleic acid can be radioactively labeled at one nucleotide position and another nucleic acid can be radioactively labeled at three nucleotide positions. The radioactive particles emitted by each nucleic acid can be detected and quantified, for example

in a scintillation counter, and displayed as the number of counts per minute (cpm). The nucleic acid labeled at three positions will emit about three times the number of radioactive particles as the nucleic acid labeled at one position and hence about three times the number of cpms will be recorded.

As used herein, the term "dendrimer" is intended to mean a branched nucleic acid. These structures are composed of layers of nucleic acid, each layer being composed of partially single-stranded heteroduplexes which are called dendrimer monomers. The outermost layer of a given dendrimer can have multiple single-stranded arms capable of hybridization with a complementary nucleic acid sequence. Dendrimer monomers have the property that sequential addition of monomers yields a three-dimensional structure composed of nucleic acid. Various configurations of nucleic acid molecules can give rise to a large number of differently shaped dendritic structures. For example, using commercially available dendrimeric synthons, a dendrimer with 1 stem and 81 branches can be synthesized. Also fork-like, comb-like and bubbled structures are possible.

As used herein, the term "nucleic acid" is intended to mean a single- or double-stranded DNA or RNA molecule including, for example, genomic DNA, cDNA and mRNA. The term is intended to include nucleic acid molecules of both synthetic and natural origin. A nucleic acid molecule of the invention can be of linear, circular or branched configuration, and can represent either the sense or antisense strand, or both, of a native nucleic acid molecule. A nucleic acid molecule of the invention can further incorporate a detectable moiety such as a radiolabel, a fluorochrome, a ferromagnetic substance, a luminescent tag or a detectable moiety such as biotin.

As used herein, the term "hybridizing" is intended to mean joining different components together. Any number of components can be joined together, for example two components can be joined together to make a duplex, three components can be joined together to make a triplex, and so on. Nucleic acids can form a hybrid or duplex, for example, by hydrogen bonding between complementary nucleotides. The formation of nucleic acid hybrids is dependent on several conditions known in the art, including temperature, salt concentration, and pH.

As used herein, "complexity" refers to the degree of repeated elements between two nucleic acids that are being hybridized together in a solution (see Anderson, M.L.M., Nucleic Acid Hybridization, Springer-Verlag, New York (1999)). When the nucleic acid molecules that are to be hybridized contain repeated core elements or homopolymeric regions, there are many possible pairing opportunities and so the hybridization proceeds quickly. When the nucleic acid molecules that are to be hybridized do not contain any repeated core elements then there is only one way that the two sequences can be hybridized and so the hybridization proceeds more slowly. Sequences that hybridize quickly are said to have a low complexity, while sequences that take longer to hybridize have a higher complexity. For example, a 40 base pair genedigit sequence made up of five direct repeats of an 8 base pair core element, can be hybridized to a 24 base pair anti-genedigit containing three repeats of the 8 base pair core repeat in three different registers (see, for example, Figure 1C). Thus the anti-genedigit can hybridize to the 40 base pair genedigit through a 24 base pair sequence that only has a complexity of an 8 base pair sequence.

The invention provides a diverse population of labels and methods for generating a large number of unique labels of about the same unit signal starting from just a small number of different labels. Enough labels are generated by this method so that each analyte in a complex mixture can be uniquely bound by a label and thus identified. These labels can be used in very small volumes which improves kinetics in the binding reaction. In addition, the design of these labels allows for improved accuracy in detection.

A large diversity of unique labels can be desirable in order to provide a unique label to each species in a complex mixture. The invention provides methods for combining different labels in pre-determined ratios to generate a large diversity of unique labels. The labels are designed in a modular fashion which allows for flexibility in the number of unique labels that are generated. For example, if a large number of modules are used, a large number of ratios of the different labels is possible which leads to a large number of unique labels. The number of labels generated can be adjusted to cover mixtures with different numbers of species.

The invention provides a diverse population of labels that contains thirty or more unique labels where each unique label is bound to a nucleic acid. A diverse population of labels is a mixture of distinct label species. This population can have as few as about thirty distinct label species or as high as 10^{17} distinct label species. The actual number of molecules of each label species can vary as long as at least one molecule of the label species is present. In addition, the invention provides a diverse population of labels that contains 40, 60, 80, 100, 120, 140, or about 150 unique labels. A portion of this population can be made up of different

individual label monomers. The invention also provides unique labels made from combinations of different labels which can increase the number of unique labels substantially.

The labels of the invention are bound to nucleic acids. In particular, the labels are bound to nucleic acids through the attachment of a label monomer to a nucleotide within a nucleic acid in a 1:1 correspondence. A nucleic acid can contain several label monomers, however each label monomer is attached directly to a nucleotide.

A label monomer can be attached to any nucleotide including both natural and non-natural nucleotides. A nucleotide contains three parts, a phosphate group, a pentose five-carbon sugar molecule, and an organic base. In RNA, the pentose is ribose and in DNA it is deoxyribose and so nucleotides for incorporation into RNA are called ribonucletides and nucleotides for incorporation into DNA are called deoxyribonucleotides. Three bases adenine, guanine, and cytosine are found in both DNA and RNA while thymine is normally found only in DNA and uracil is normally found only in RNA. Nucleotides can have one, two or three attached phosphate groups and are sometimes referred to as nucleoside phosphates. Nucleotides can contain modified nucleosides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'O-methyl ribosyl, 2'O-methoxyethyl ribosyl, 2'fluoro ribosyl, 2'amino ribosyl, and the like). An example of non-natural bases that are used in the art are isocytidine and isoguanine.

A label monomer as used herein is intended to mean an individual measurable moiety, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent

marker, biotin, or other moiety known in the art that is measurable by analytical methods. A label monomer can be attached to a nucleotide using methods well known in the art and exemplified herein.

Radioisotopes are an example of label monomers that can be utilized by the invention. Several radioisotopes can be used as label monomers for labeling nucleotides including, for example, ^{32}P, ^{33}P, ^{35}S, ^{3}H, and ^{125}I. These radioisotopes have different half-lives, types of decay, and levels of energy which can be tailored to match the needs of a particular experiment. For example, ^{3}H is a low energy emitter which results in low background levels, however this low energy also results in long time periods for autoradiography. Radioactively labeled ribonucleotides and deoxyribonucleotides are commercially available. Nucleotides are available that are radioactively labeled at the first, or α, phosphate group, or the third, or γ, phosphate group. For example, both [α-^{32}P]dATP and [γ-^{32}P]dATP are commercially available. In addition, different specific activities for radioactively labeled nucleotides are also available commercially and can be tailored for different experiments.

Another example of label monomers that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, and Texas Red. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum. Also different formulations of the same fluorophore have been produced for different applications. For example, fluorescein, can be used in its isothiocynanate form (FITC), as mixed

isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or as isomeric dichlorotriazine forms of fluorescein (DTAF). These labels are chemically distinct, but all emit light with a peak between 515-520 nm. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein. In addition, different formulations or related dyes have been developed around other fluorophores that emit light in other parts of the spectrum.

Amine-reactive and thiol-reactive fluorophores are available and used for labeling nucleotides and biomolecules. Generally, nucleotides are fluorescently labeled during chemical synthesis, for example, incorporation of amines or thiols during nucleotide synthesis permit addition of fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum.

Fluorescent dyes that can be bound directly to nucleotides can also be utilized as label monomers. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These fluorecently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-ddGTP and ROX-ddUTP, are commercially available.

Non-radioactive and non-fluorescent label monomers are also available. For example, biotin can be attached directly to nucleotides and detected by specific and high affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin labeled nucleotides can also similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

Very small particles, termed nanoparticles, also can be used as label monomers to label nucleic acids. These particles range from 1 - 1000nm in size and include diverse chemical structures such as gold and silver particles and quantum dots.

When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle. Four to five different particles in close proximity will each scatter monochromatic light which when superimposed will give a specific, unique color. The particles are being manufactured by companies such as Genicon Sciences. Derivatized silver or gold particles can be attached to a broad array of molecular probe molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivitized to allow attachment to a nucleotide.

Another type of nanoparticle that can be used as a label monomer are quantum dots. Quantum dots are fluorescing crystals 1-5nm in diameter that are excitable by a large range of wavelengths of light. These crystals

emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties. These particles have been used in the semi-conductor industry for several years, but are just now being applied to molecular biology applications.

Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to desired size and/or emission wavelengths. Use of quantum dots for labeling particles, in the context of the present invention, is new, but is old in the art of semiconductors. Two examples of earlier references in which quantum dots are embedded within intrinsic silicon epitaxial layers of semiconductor light emitting/detecting devices are United States Patent Nos. 5,293,050 and 5,354,707 to Chapple-Sokol, et al.

Due to their very small size the quantum dots can be coupled into oligonucleotides directly without affecting the solubility or use of the oligonucleotide. The invention requires that only one oligonucleotide molecule is coupled to each nanoparticle. To synthesize an oligonucleotide-nanoparticle complex in a 1:1 ratio by conventional batch chemistry, both the oligonucleotide and the nanoparticle require a single reactive group of different kinds that can be reacted with each other. For example, if an oligonucleotide has an amino group and a nanoparticle has an aldehyde group, these groups can

react to form a Schiff base. An oligonucleotide can be derivitized to attach a single amino or other functional group using chemistry well known in the art. However, when a nanoparticle is derivatized, it is covered with a chemical reagent which results in coating the entire surface of the nanoparticle with several functional groups.

The invention provides a method of coupling one oligonucleotide to one nanoparticle by chemically coupling the oligonucleotide on a solid surface such as the glass support used for the oligonucleotide synthesis. For example, commercially available resins for oligonucleotide synthesis such as long chain alkylamino controlled pore glass (lcaa CPG) can be used. Alternatively, a flat surface such as a derivitized microscope slide can be used. The surface density of the nascent oligonucleotide chains should be lower than the diameter of the nanoparticle. This can be achieved by either choosing a glass support with low surface density of the reactive groups, or by using diluted reagent for the first step of the oligonucleotide synthesis so that the surface is not saturated. Another point of consideration when using the standard glass matrices for oligonucleotide synthesis is to use a pore diameter higher than the nanoparticle diameter to ensure the flow of the reagents. For example, an oligonucleotide can be synthesized on a diluted basis relative to the solid support, for example one tenth of a normal synthesis, to ensure good spacing of the oligonucleotides on the glass support. After the oligonucleotide is synthesized with a reactive functional group, for example, an amino group, derivitized nanoparticles are passed over the glass support to react with the oligonucleotides. A sufficiently large pore size of the glass support can be chosen to prevent clogging with nanoparticles. For

example, a pore size of about 200 nm can be used. After the reaction is complete, un-reacted groups on the nanoparticle can be blocked and the complexes can be uncoupled from the glass support.

The labels of the invention are bound to nucleic acids through nucleotides in the nucleic acid. A nucleotide can be attached to a label monomer first and then the label monomer:nucleotide can be incorporated into a nucleic acid, or an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid.

A label monomer can be attached to a nucleotide using a variety of methods well known in the art and described herein. For example, the label monomer can be directly attached to the nucleotide in a 1:1 correspondence by incorporation of a radioactive phosphate into the phosphate backbone of the nucleotide. Also, for example, a general method for labeling phosphates with a fluorescent label that employs an imidazole derivative prepared from a BODIPY FL hydrazide has been reported (Wang and Giese, Anal. Chem. 65: 3518 (1993).

Depending on the labeling moiety used, it can be desirable to derivitize or chemically modify a nucleotide in order to bind the label monomer. These methods and chemistries are known in the art. In addition, a linker can be used to attach a label monomer to a nucleotide in a 1:1 correspondence. For example, a fluorescently labeled nucleotide such as fluorescein-12-dUTP can have a fluorophore monomer attached via a four-atom aminoalkynyl group to the dUTP molecule.

These nucleotides attached to label monomers can be incorporated into a nucleic acid using several methods for labeling nucleic acids well known in the art. For example, enzymes such as DNA or RNA polymerases, Taq polymerases, terminal deoxynucleotidyl transferases, or reverse transcriptases can be used to incorporate labeled nucleotides into nucleic acids.

Labeled nucleotides can be incorporated into nucleic acids, for example, by nick translation. In this procedure DNAse I is used to create single-strand nicks in double stranded DNA and then the 5' to 3' exonuclease and 5' to 3' polymerase actions of E. coli DNA polymerase I are used to remove stretches of single stranded DNA starting at the nicks and replace them with new strands made by incorporation of labeled nucleotides. Nick translation can utilize any labeled nucleotide including radioactively labeled nucleotides and biotinylated or digoxigenin labeled nucleotides. In a similar way T4 DNA polymerase can be used to incorporate labeled nucleotides. In addition, labeled nucleotides can be incorporated into nucleic acids using the polymerase chain reaction (PCR) and Taq polymerases. The degree of labeling can be controlled by including one, or up to all four labeled nucleotides. In addition, the degree of labeling can be controlled by increasing or decreasing the concentration of the labeled nucleotide(s).

Other methods for labeling nucleic acids include generating single-stranded cDNA from RNA by using a reverse transcriptase in the presence of labeled nucleotides. In addition, DNA can be cloned into a vector with SP6 or T7 polymerase sites. Transcription in the presence of SP6 or T7 RNA polymerase and labeled nucleotides results in a labeled RNA transcript. The transcript can be labeled to different degrees by

including one or more labeled nucleotides. In addition, several nucleotides within a nucleic acid can be labeled, for example, by cloning DNA into a bacteriophage M13 based vector. Then the Klenow fragment of DNA polymerase I and the M13 universal probe primer can be used to synthesize the complementary stand with incorporation of labeled nucleotides.

Several methods are described above for incorporation of labeled nucleotides into newly synthesized nucleic acids. Existing nucleic acids can also be labeled using several methods known in the art. For example, RNA or DNA can be end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. This kinase can be used to transfer the radioactive phosphate of ATP to a free 5' OH group in either DNA or RNA. The enzyme also has a phosphatase activity and so two reactions are possible. In the forward reaction, the enzyme catalyzes phosphorylation following removal of 5' terminal phosphates with alkaline phosphatase (or other phosphatase). In the exchange reaction, the kinase catalyzes the exchange of an existing 5' phosphate with the third or γ phosphate of ATP. The latter reaction is carried out in the presence of excess ATP and ADP for efficient phosphorylation. Using this method the radioactive phosphate of ATP is transfered to the end of the nucleic acid molecule.

Nucleic acids can also be labeled with terminal deoxynucleotidyl transferease which adds labeled nucleotides onto the 3' end of DNA fragments. Both single and double-stranded DNAs are substrates for this enzyme. The large (Klenow) fragment of E. coli DNA polymerase I can also be used to label the ends of nucleic acids. Since this enzyme has a 5' to 3'

polymerase activity it can be use to "fill in" the 3' ends of DNA fragments opposite of 5' extensions or overhangs with labeled nucleotides. End-labeling of nucleic acids using polynucleotide kinase or terminal deoxynucleotidyl transferease results in the incorporation of one label per nucleic acid. The "fill in" reaction can be used to label the nucleic acid at one nucleotide per nucleic acid or at more than one nucleotide per nucleic acid.

In addition, nucleic acids can be labeled by modification of nucleotides within the nucleic acid. For example, cytidine residues in DNA and RNA can be modified by reaction with sodium bisulfite to form sulfonate intermediates that can then be directly coupled to hydrazides or aliphatic amines. Virtually any of the fluorescent, biotin or other hydrazides or aliphatic amines can be used in this reaction. The bisulfite-activated cytidylic acid can also be coupled to aliphatic diamines such as ethylenediamine. The amine-modified DNA or RNA can then be modified with any of the amine-reactive dyes. In addition, phosphate groups can be targeted in nucleic acids for labeling. Although phosphate groups of nucleotides are not very reactive in aqueous solution, their terminal phosphate groups can react with carbodiimides and similar reagents in combination with nucleophiles to yield labeled phsophodiesters, phosphoramidates and phosphorothioates. For example, DNA can be reacted quantitatively with carbonyl diimidazole and a diamine such as ethylenediamine to yield a phosphoramidate that has a free primary amine and that this amine can then be modified with amino-reactive reagents. Fluorescent or biotinylated amines have been coupled to the 5' phosphate of tRNA using dithiodipyridine and triphenylphosphine.

The bond between labels and nucleic acids can be covalent bonds or non-covalent bonds that are stable to hybridization and washing conditions. The labels can be bound to a nucleic acid in a sequence specific manner, for example by the incorporation of a labeled nucleotide into DNA that has been digested by a restriction enzyme. Alternatively the labels can be bound to a nucleic acid in a non-sequence specific manner, for example by the incorporation of a label onto the terminal phosphate of a nucleic acid using [γ-^{32}P]ATP and T4 polynucleotide kinase.

Several types of nucleic acids can be utilized with this invention including a single- or double-stranded DNA or RNA molecule which can include, for example, genomic DNA, cDNA and mRNA. Nucleic acid molecules of both synthetic and natural origin can be used. A nucleic acid molecule of the invention can be of linear, circular or branched configuration, and can represent either the sense or antisense strand, or both, of a native nucleic acid molecule. Nucleotide bases not found in nature such as isocytidine and isoguanine can be incorporated into the nucleic acid.

Labels can bind to nucleic acids in a number of diverse ways. For example, a particular label monomer can be bound to a nucleic acid at only one position in the nucleic acid or at many positions in the nucleic acid. In addition, a particular label monomer can be bound to a nucleic acid and one or more other label monomers can also bound to the same nucleic acid. In this case the label can contain a mixture of two or more different labels. Furthermore, nucleic acids labeled with any or all of these combinations can be bound to another nucleic acid through hybridization.

Additional diversity is introduced when the nucleic acid is branched. One example of a branched nucleic acid is a dendrimer. Dendrimers are composed of layers of nucleic acid, each layer being composed of partially single-stranded heteroduplexes yielding a three-dimensional structure composed of nucleic acid. Various configurations of nucleic acid molecules can give rise to a large number of differently shaped dendritic structures including, for example, a dendrimer with 1 stem and 81 branches, or dendrimers with fork-like, comb-like or bubbled structures. The outermost layer of a given dendrimer can have multiple single-stranded arms capable of hybridization with a complementary nucleic acid sequence. Due to the relatively large size of nucleic acid molecules, nucleic acid dendrimers can contain numerous labels with limited steric hindrance. Use of dendrimers can multiply the signal generated by a labeled nucleic acid by a pre-determined factor equal to the number of branches.

Several unique combinations of labels can be formed using branched nucleic acids. For example, by using different chemical protective groups, one label monomer can bind to one branch while one or more other label monomers bind to other branches. Furthermore, labeled nucleic acids can be attached in various combinations to the branches of a dendrimer through hybridization.

The invention provides a diverse population of 30, or about 40, 60, 80, 100, 120, 140, or 150 unique labels bound to a nucleic acid. Part of this population can be made up of different individual label monomers. The invention also provides unique labels made from a combination of two or more different labels. This can increase the number of unique labels substantially.

A unique label is a label that generates a signal that is distinguishable from other labels in the same mixture. Therefore, designation as a unique label is dependent upon the sensitivity of the detection equipment that is used. For example, where fluorescent or nanoparticle labels are used a CCD camera can be used to detect the labels. The sensitivity of this equipment depends on the manufacturer, model, and design of the equipment. In addition, several parameters can be set by the user in order to achieve maximum sensitivity. For example, the use of different filter sets can increase the sensitivity of detection for certain experiments.

The ability to distinguish different labels also depends on the particular properties of the label. For example, some fluorophores emit light within a broad peak or range of wavelengths while other fluorophores emit light within a narrow peak. Fluorophores that emit light within a broad peak can obscure neighboring peaks. In addition, the shapes of the emission peaks from different fluorophores varies. For example, some fluorophores have a peak that has a sharp ascent, but a broad trailing end that can obscure neighboring peaks. If two peaks cannot be distinguished then the two labels associated with those peaks cannot be considered unique. Where fluorescent labels are used, the labels would emit fluorescent light at distinct, essentially non-overlapping wavelengths distant from each other by at least 1nm, 5nm, 10nm, 15nm, 20nm, 25nm, and preferably 30nm, 35nm, 40nm, 45nm, and more preferably by at least 50nm. For example, the emission peak of dye #1 could be 585nm, and the peak emission of dye #2 could be 630nm.

A unique label is a label that generates a signal that is distinguishable from other labels in the same mixture. A unique label, therefore, is dependent on the

other labels that are included in the mixture. For example, fluorescein which has an emission spectra peak at 518nm and rhodamine red with an emission spectra peak of 590nm have clearly distinct emission peaks and hence both are considered to be unique labels when included together in a mixture. However, if another fluorophore is added that has an emission peak that overlaps with fluorescein or rhodamine red, these labels would not be unique. For example, if Oregon green (emission spectra peak 522nm) is added to the above mixture the emission peak of fluorescein and Oregon Green can overlap so much that, depending on the detection equipment used, they cannot be clearly distinguished from each other and hence neither fluorescein nor Oregon Green can act as unique labels in that particular mixture.

Often several formulations of the same label or related labels are synthesized for use in different applications, and while these labels have different chemical properties, they are not distinct in terms of detection. For example, fluorescein, can be used in its isothiocynanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or isomeric dichlorotriazine forms of fluorescein (DTAF). These labels are chemically distinct, but all emit light with a peak between 515-520 nm and hence would overlap so as to appear identical on most currently available detection equipment. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their

emission peaks and so also serve functionally as substitutes for fluorescein.

A limited number of different label monomers are known that can be used together in a mixture and still provide unique signals. For example, five analytes can be distinctly labeled using the BODIPY fluorophore set from Molecular Probes (Eugene, OR). These fluorophores have the following distinct emission peaks: BODIPY FL(513), BODIPY R6G (550), BODIPY TMR (574), BODIPY 581/591 (592) and BODIPY TR (617). In order to obtain clear results using currently available detection equipment, the number of different fluorophores that can be used is less than thirty.

Unique labels of the invention also can be generated by combining two or more different label monomers to make a new label. The signal from the resulting label must be distinguishable from the signals of other labels used in the same experiment in order to be a unique label. For example, a nucleic acid labeled with both fluorescein and rhodamine will emit light at a different wavelength than a nucleic acid labeled with either fluorescein or rhodamine alone.

In this invention, various ratios of different label monomers bound to nucleic acids can be combined to generate a diverse population of unique labels that can include up to 10^{17} or more unique labels. For example, a nucleic acid labeled with two fluorescein labeled nucleotides and three rhodamine labeled nucleotides will emit light at a different wavelength compared to a nucleic acid labeled with three fluorescein nucleotides and two rhodamine nucleotides. In another example, a nucleic acid could be labeled with different ratios of three or more label monomer:nucleotides which greatly

increases the variety of unique labels that can be generated.

The signal generated by each of the label monomers bound to the nucleotides can be normalized to have about the same unit signal. For example, if fluorescent monomer A is known to emit a different quanta of light as fluorescent monomer B, the signal from unique labels that contain a pre-determined number of these monomers can be normalized based on the known signal properties of the label monomers and the number of each monomer present in the unique label. Different numbers of labels can be used in the invention and so different multiples of the same unit signal can be utilized by the invention. For example, a nucleic acid can be labeled with two fluorophores and another nucleic acid can be labeled with six fluorophores. The second nucleic acid will have three times the signal of the first nucleic acid. Since the number of label monomers attached to each nucleic acid is known, the signals from the labeled nucleic acids can be normalized based on the number of label monomers present. For example, the signal from the nucleic acid with six fluorphores can be divided by three which normalized the signal relative to the signal from the nucleic acid with two fluorophores.

This method of producing labels results in important advantages over existing techniques such as microarray formats. Since each analyte is uniquely identified by a label with about the same unit signal, it allows the labels to be directly counted resulting in a digital read-out of each molecular species in a mixture. In contrast, microarray data must undergo several intermediate transformations to quantitate the number of molecules which results in a less precise analog output. In addition, using a diversity of labels with the same

unit signal allows the detection method to operate in a narrow dynamic range resulting in improved sensitivity of the system since the trade-off between sensitivity and dynamic range is avoided.

The invention provides a diverse population of uniquely labeled probes. This population of probes contains about 30 or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid. In addition, the invention provides a diverse population of uniquely labeled probes containing a diversity of 50, 100, 200, 500, 1,000, 2,000, 5,000, $1x10^4$, $3x10^4$ and about $1x10^5$ or more different labels. As described above, these unique labels can contain a mixture of two or more different labels and comprise about the same unit signal or multiple thereof.

The invention provides a diverse population of uniquely labeled probes in which a target specific nucleic acid contains a nucleic acid bound to a unique label. In addition, the invention provides a diverse population of uniquely labeled probes containing two attached populations of nucleic acids, one population of nucleic acids containing thirty or more target specific nucleic acid probes, and a second population of nucleic acids containing a nucleic acid bound by a unique label.

A target specific probe is intended to mean an agent that binds to the target analyte selectively. This agent will bind with preferential affinity toward the target while showing little to no detectable cross-reactivity toward other molecules.

The target analyte can be any type of macromolecule, including a nucleic acid, a protein or even a small molecule drug. For example, a target can be a nucleic

acid that is recognized and bound specifically by a complementary nucleic acid including for example, an oligonucleotide or a PCR product, or a non-natural nucleic acid such as a locked nucleic acid (LNA) or a peptide nucleic acid (PNA). In addition, a target can be a peptide that is bound by a nucleic acid. For example, a DNA binding domain of a transcription factor can bind specifically to a particular nucleic acid sequence. Another example of a peptide that can be bound by a nucleic acid is a peptide that can be bound by an aptamer. Aptamers are nucleic acid sequences that have three dimensional structures capable of binding small molecular targets including metal ions, organic dyes, drugs, amino acids, co-factors, aminoglycosides, antibiotics, nucleotide base analogs, nucleotides and peptides (Jayasena, S.D., Clinical Chemistry 45:9, 1628-1650, (1999)) incorporated herein by reference. Further, a target can be a peptide that is bound by another peptide or an antibody or antibody fragment. The binding peptide or antibody can be linked to a nucleic acid, for example, by the use of known chemistries including chemical and UV cross-linking agents. In addition, a peptide can be linked to a nucleic acid through the use of an aptamer that specifically binds the peptide. Other nucleic acids can be directly attached to the aptamer or attached through the use of hybridization. A target molecule can even be a small molecule that can be bound by an aptamer or a peptide ligand binding domain.

The invention provides a method of producing a population of labels consisting of synthesizing a population of nucleic acids bound to a predetermined ratio of at least two different labels. The method involves incorporating labeled nucleotides into a repeated nucleic acid structure using a DNA polymerase. The repeated nucleic acid structures can be designed to

allow incorporation of a pre-determined ratio of labels. Using this method several unique labels can be generated from a small number of starting labels.

A specific example of this method where ten unique labels are made from two different labels is provided in Example 1. Briefly, ten unique templates of a 220 base pair single-stranded DNA are synthesized. The templates consist of a pre-determined ratio of the following 20 base pair repeats where n+m=11: 5' (ACTCTCTCTCTCTCTCTCTC)n(GCTCTCTCTCTCTCTCTCTC)m 3' (SEQ ID NO:1). The second strand is synthesized using the primer GAGAGAGAGA (SEQ ID NO:2), Klenow polymerase, DNA ligase, dGTP, dATP, and dCTP and dUTP each labeled with a different fluorophore. The labeled nucleotides will be incorporated into the DNA in a unique ratio determined by the ratio of the two repeats. In this example, the end result is ten uniquely labeled nucleic acids where the set ratio of the two fluorophores is 1:10, 2:9, 3:8, 4:7, 5:6, 6:5, 7:4, 8:3, 9:2, and 10:1.

In Example 1, two different labels resulted in ten unique labels. Using the same protocol, three different labels would result in 30 unique labels, four different labels would result in 60 unique labels, five different labels would result in 100 unique labels, and so on. Several variations of the method will be apparent to one skilled in the art. For example, the number of repeats can be changed to be less than or greater than ten. An increased number of repeats will increase the number of unique ratios possible. This will result in an increase in the number of unique labels that can be generated from the same number of starting different labels. Also in Example 1, the ratio between the two fluorophores can include 0:11 and 11:0 which results in two additional labels that contain one fluorophore or the other.

One skilled in the art will recognize that the sequence of the templates can differ from that shown above. For example, the repeat sequence in the template can be $(GA)_n$ instead of $(CT)_n$. In addition, the repeat sequence could be a single nucleotide homopolymer such as $(A)_n$. With a homopolymer template, three labeled nucleotides can be incorporated in different ratios thus increasing the number of unique labels that can be generated. It is possible using the claimed method to generate a large number of unique labels including 40, 60, 80, 100, 120, 140, 150, 200, 500, 2,000, 5,000, $1x10^4$, $3x10^4$, $1x10^5$ or more labels. Again as described above, these unique labels can comprise about the same unit signal or multiple thereof.

Another possible modifications of the method is to change the length of the repeat, for example, to less than or greater than 20 base pairs. The repeat serves to separate the labeled nucleotides and therefore to decrease possible interference between the labels. In the case of a fluorescent label, this can decrease quenching between fluorophores. In addition, the protocol for incorporating the labeled nucleotides into the DNA can be modified as would be clear to one skilled in the art and as described herein.

In one embodiment, the labeled DNA described above can be attached to a dendrimer. Oligonucleotide tags can be synthesized at the branches of the dendrimer to allow binding of the labeled DNA. For example, a linker can be attached to the labeled DNA described above that corresponds to an oligonucleotide tag on the branches of the dendrimer. Several types of linkers are known to one skilled in the art. For example, a restriction enzyme linker can be attached to the labeled DNA. These linkers are double-stranded oligonucleotides that contain the

recognition sequence of a particular restriction enzyme. These linkers can be ligated onto double-stranded DNA using a DNA ligase and digested using the appropriate restriction enzyme. The result is an overhanging single stranded sequence that is available to hybridize to another nucleic acid.

The labeled DNA described above can be directly attached to a target specific probe. In addition, the labeled DNA can be indirectly attached to a target specific probe, for example, through use of a bridging nucleic acid. One or more of these labels can be attached to each target specific probe. Binding of a uniquely labeled target specific probe to a target analyte results in the unique tagging of that analyte. This tagging allows identification of the target analyte from a mixture of analytes.

The number of unique labels can be further increased by combining the unique labels described above in different combinations. The invention provides a method of attaching a label to a nucleic acid probe, comprising hybridizing a nucleic acid probe having a genedigit to an anti-genedigit having a label. The anti-genedigits that contain the labels described above are hybridized to genedigits. Genedigits can be linked together in unique combinations creating an even larger number of unique labels. The modular design of the genedigits allows for flexibility in the number of unique labels that are generated. For example, if a large number of genedigit modules are used, a large number of unique templates will be available for attachment of unique labels. The number of templates generated can be adjusted to cover the number of species in the mixture.

A genedigit can be a region of pre-determined nucleotide or amino acid sequence that serves as an attachment point for a label. The genedigit can have any sequence including, for example, a single unique sequence or a sequence containing repeated core elements. However, each genedigit has a unique sequence which differentiates it from other genedigits. When added to a complex mixture of targets, a nucleic acid genedigit can contain non-natural bases such as isocytidine and isoguanine which can reduce hybridization to naturally occurring target sequences. The sequence, length of a core element, and number of repeated core elements can be varied according to the particular requirements of an experiment and will be clear to one skilled in the art. For example, a nucleic acid core element can be between about 5 and 12 base pairs in length and the core element can be repeated once to about ten times.

Different genedigits can be synthesized that have a core unit that differs from the core units of other genedigits, for example, by at least two bases. By combining these unique genedigits in various combinations, a highly diverse number of structures can be synthesized. For example, 50 genedigits with unique sequences can be synthesized and split into five groups containing ten genedigits in each group. The genedigits of each group can be synthesized to have a short tag on each end. One genedigit from each group is then linked together using an adapter oligonucleotide that is complimentary to the tags that are common for each group. In this example, $1x10^5$ (10x10x10x10x10) unique combinations are possible.

Genedigits serve as attachment points for the unique labels described above. Since genedigits can be linked together in unique combinations this greatly increases

the number of unique labels. By linking together genedigits, a large number of unique labels can be generated including 200, 500, 2,000, 5,000, $1x10^4$, $3x10^4$, $1x10^5$ or more unique labels.

The invention provides a method of attaching a label to a nucleic acid probe, comprising hybridizing a nucleic acid probe having a genedigit to an anti-genedigit having a label where the genedigit comprises a set of three or more repeat sequences, and the anti-genedigit comprises a cognate set of at least two complimentary repeat sequences where the anti-genedigit specifically hybridizes to the genedigit through a sequence having a complexity less than the number of hybridized base pairs.

An example of a nucleic acid genedigit is given in Example 2. Briefly, in this example an eight base pair core element containing the non-natural bases isocytidine and isoguanine is repeated five times. This results in a 40 base pair genedigit. In this example the anti-genedigit consists of a sequence that is complimentary to three of the five core elements in the genedigit. An anti-genedigit can contain the same number, or a lesser number, of repeat sequences compared to the genedigit as long as the anti-genedigit is able to specifically bind to the genedigit. In this example, the anti-genedigit is a 24 base pair sequence that can bind to the 40 base pair genedigit in three different registers (see Figure 1C).

The anti-genedigit in this specific example is a 24 base pair sequence that only has the complexity of an eight base pair sequence. Complexity refers to the degree of repeated elements between two nucleic acids that are being hybridized together in a solution. When the nucleic acid molecules that are to be hybridized contain repeated core elements or homopolymeric regions,

there are many possible pairing opportunities and so the hybridization proceeds quickly. When the nucleic acid molecules that are to be hybridized do not contain any repeated core elements then there is only one way that the two sequences can be hybridized and so the hybridization proceeds more slowly. Sequences that hybridize quickly are said to have a low complexity, while sequences that take longer to hybridize have a higher complexity. In this example, a 40 base pair genedigit sequence made up of five direct repeats of an 8 base pair core element, can be hybridized to a 24 base pair anti-genedigit containing three repeats of the 8 base pair core repeat in three different registers. Thus the anti-genedigit can hybridize to the 40 base pair genedigit through a 24 base pair sequence that only has a complexity of an 8 base pair sequence. The advantage of this method is that hybridization will proceed more quickly and efficiently.

In the specific example above, an 8 base pair core element was describe; however, a core element can be more or less than 8 base pairs. For example, a core element can be between 5 and 12 base pairs. A change in the core element of a repeat will correspondingly change complexity. For example, if a core element is between 5 and 12 base pairs, complexity will be between 5 and 12. In addition, an anti-genedigit can bind to a genedigit using all or part of its sequence. In the example above, a 24 base pair anti-genedigit can bind to the genedigit with less than 24 bases, for example, 15 to 23 base pairs.

The anti-genedigit serves as a connector between the genedigit and the label. The genedigit is bound by the anti-genedigit and the anti-genedigit is bound to a label directly or indirectly by being bound to a dendrimer that

has a label(s) attached. The anti-genedigit also contains a linker sequence that allows attachment to a dendrimer. For example, an oligonucleotide tag can be synthesized at the stem of a dendrimer that is complementary to the linker sequence on the anti-genedigit. As described above, several types of linkers can be used. In this way the repeat sequences of the anti-genedigit are free to hybridize with their corresponding genedigit resulting in the specific attachment of a label to a genedigit.

Genedigits and anti-genedigits can be comprised of nucleic acids, including aptamers, as well as macromolecules other than nucleic acids. For example, a genedigit can be comprised of an amino acid sequence that is bound by an anti-genedigit that is a nucleic acid that binds specifically to the amino acid sequence in the genedigit, or the anti-genedigit can be an amino acid sequence, including an antibody or antibody fragment, that specifically binds to the amino acid sequence in the genedigit.

The invention provides for a method whereby a target specific probe can be attached to one or more genedigits to form a "specifier." The genedigits can be directly linked or can be attached to the target specific region using an intervening or adapting sequence. As described above, the target specific area can be a nucleic acid, including an aptamer, or the target specific area can be an amino acid sequence, including an antibody or antibody fragment. The target specific area is designed to specifically bind an analyte in a mixture. In this way an analyte can be labeled with a unique label.

A specifier can contain one to several genedigits. For example, a specifier can contain four or five or more

gendigits. The number of genedigits in a specifier will determine the number of unique labels that are available to bind to analytes in a mixture. Therefore, in order to label each analyte in a complex mixture, a large population of specifiers can be synthesized that contain several combinations of genedigits. Alternatively, to label one or a small number of analytes in mixture, a specifier or specifiers can be synthesized that contain one or a few genedigits. In addition, a specifier can contain a common tag such as a biotin tag. These tags can facilitate synthesis and purification of specifiers.

The invention additionally provides a method of producing a population of target specific nucleic acid probes. The method consists of producing a first population of two or more target specific probes each having at least one genedigit, where the genedigit contains a set of three or more repeated sequences then producing a second population of nucleic acids having an anti-genedigit with at least two complimentary repeated sequences, and finally hybridizing the first and second populations of nucleic acids to produce a population of target specific probes attached to an anti-genedigit where the anti-genedigit hybridizes to the genedigit through a sequence having a complexity less than the number of hybridized base pairs.

The invention also provides a method for producing a population of uniquely labeled nucleic acid probes. The method consists of synthesizing a population of target specific nucleic acid probes each having a different specifier; then synthesizing a corresponding population of anti-genedigits each having a unique label, and finally hybridizing the populations of target nucleic acid probes to the anti-genedigits, to produce a

population where each of the target specific probes is uniquely labeled.

The invention further provides a method for detecting a nucleic acid analyte, by contacting a mixture of nucleic acid analytes with a population of target specific probes each attached to a unique label under conditions sufficient for hybridization of the probes to the target and measuring the resulting signal from one or more of the target specific probes hybridized to an analyte where the signal uniquely identifies the analyte.

The nucleic acid analyte can contain any type of nucleic acid, including for example, an RNA population or a population of cDNA copies. The invention provides for at least one target specific probe for each analyte in a mixture. The invention also provides for a target specific probe that contains a nucleic acid bound to a unique label. Furthermore, the invention provides two attached populations of nucleic acids, one population of nucleic acids containing a plurality of target specific nucleic acid probes, and a second population of nucleic acids containing a nucleic acid bound by a unique label. When the target specific probes are attached to unique labels, this allows for the unique identification of the target analytes.

The methods of the invention are advantageous because hybridizations can be performed in solution in a small volume (0.01 -2.0 μl) which ensures high concentration of the nucleic acids which will drive the hybridization rate. Two different types of hybridizations are utilized by the methods of invention. The first hybridization is between a complex mixture of analytes and the specifiers and the second type of hybridization is between the specifiers and the labels.

In the first type of hybridization between a complex mixture of analytes and a population of specifiers, the specifers are in excess. For example, the specifiers can be in 10 to about 100 fold excess over the analytes in the complex mixture. The kinetics of this reaction can be described by the following equation:

$$t_{1/2} = \frac{N \ln 2}{3.5 \times 10^5 \times L^{0.5} \times C_0}$$

where N is the complexity of the probe (the specifiers), L is the length, C_0 is the concentration of the probe, and $t_{1/2}$ is the time required for 50% completion of the reaction. Using this equation it is clear that higher concentrations of probe will result in shorter time required for 50% completion of the reaction. The sample in this hybridization can be either RNA or DNA. If the sample is not poly A RNA then it must first be tagged by some method, for example, platinum-digoxygenin. After the hybridization is complete, the duplexes containing the analyte and specifier are separated using the poly A, or the digoxygenin tag as a handle and washed from the excess specifier. The washed material is then used for the second hybridization.

In the second hybridization the isolated specifiers are mixed with labels. The labels are in excess over the specifiers and the kinetics follow the same kinetics as shown above for the first hybridization, and it is performed at an even faster rate because the complexity, N, is very low. The hybridized complexes containing specifiers and labels can be isolated using a biotin tag on the specifiers as a handle and processed further for detection.

The formation of macromolecule hybrids is dependent on several conditions known in the art, including

temperature, salt concentration, and pH. Different conditions for the formation of nucleic acid hybrids are well known in the art and can be found in, for example, Hames and Higgins, Nucleic Acid Hybridisation: A Practical Approach, IRL Press, Oxford (1991). In addition, conditions for nucleic acid-protein and protein-protein interactions are well known in the art and can be found in, for example, Current Protocols in Molecular Biology (ed. Ausubel et al., Greene Publ., New York 1989) which is incorporated herein by reference.

The "specifier-label" complexes can be separated from one another, for example, by spreading them on a two-dimensional surface such as glass, or by splitting them in liquid droplets in a flow cytometer. In this example, visualization can be achieved either by scanning the 2-D surface or by flow cytometry. Co-localization of specific labels will determine the identity of the particular specifier, which will determine the identity of the particular analyte that the specifier bound initially.

The "specifier-label" complexes can be detected by various devices including but not limited to visual inspection, digital cameras, video cameras, photographic film, or the use of current instruments such as laser scanning devices, fluorometers, luminometers, photodiodes, quantum counters, plate readers, epifluorescence microscopes, scanning microscopes, confocal microscopes, or by other means for amplifying the signal such as a photomultiplier tube or other light detector capable of detecting the presence, location, intensity, excitation and emission spectra, fluorescence polarization, fluorescence lifetime and other physical properties of the fluorescent signal. Non-fluorescent signals can be detected using a Geiger counter,

scintillation counters, chemiluminescence, enzyme assays, and autoradiography.

An example of a detection method that can be utilized by the invention is a CCD (charge-coupled device) camera. Use of this device requires that the specifier-label complexes are first spread on a glass slide. Since all complexes will have approximately the same total intensity this simplifies detection, because the camera can be adjusted for maximum sensitivity (maximum gain) and minimum dynamic range.

For example, if fluorescent dendrimers are used for labeling the sample can be excited at the corresponding absorption maxima and scanned at the corresponding emission maxima for each of the, for example, 8-10 fluorofors used. A benefit of CCD cameras is that they have very wide range of detection which allows for a choice of fluorofors with distant emission peaks. Thinned layer CCD cameras can detect from the soft X-ray to the near infrared spectrum.

Alternatively if nanoparticles labels are used the sample can be illuminated by angled white light and detection occurs at a few wavelengths. The number of wavelengths will depend on the quality of the camera used and its sensitivity and linearity. Cameras of good quality can reproducible distinguish millions of colors with only the three standard filters (red, green, and blue).

In the case of unique labels that contain more than one label monomer, the overall signal generated by the unique label can be determined. For example, an overall signal for a unique label containing several fluorophores can be determined using a spectrometer. In addition to

detecting an overall signal generated from the unique label, the invention provides for the determination of the spatial location of starting labels (which can be label monomers or combinations of label monomers) within a unique label. For example, the labels within a unique label can be separated from one another using a flow-stretch technique (Henegariu et al., Biotechniques 31:246-250 (2001)) a receding meniscus technique (Yokota et al., Nuc. Acids Res. 25:1064-1070 (1997)) or an electrostretching technique (Matsuura et al., Nuc. Acids Res. 29:E79 (2001)).

The use of flow-stretching, receding meniscus, or electro-stretching techniques allows for the separation of the labels within a unique label so that one can determine spatially where a particular label is positioned in the unique label (Example V, VI, VII and VIII). Therefore, unique labels that have the same combination of starting labels and the same overall signal can be differentiated from one another based on the location of those labels within the unique label. This ability to locate the position of a label within a unique label allows for the position of the starting label to be used as a distinguishing characteristic when generating a set of unique labels. Hence, a complex set of unique labels can be generated using the same combination of starting labels by varying the position of the labels within a unique label.

An observed signal can be modified using methods known in the art. For example, an observed signal can include subtraction of non-specific noise. An observed signal can also include, for example, treatment of the measured quantity by routine data analysis and statistical procedures which allow meaningful comparison and analysis of the observed values. Such procedures

include, for example, normalization for direct comparison of values having different scales, and filtering for removal of aberrant or artifactual values.

In the first type of hybridization described above between a complex mixture of analytes and a population of specifiers, the specifers were in excess. In contrast, hybridization on standard microarrays occurs under conditions where the analyte is in excess compared to the labeled probe. In addition, on microarrays only a small fraction of this probe will actually hybridize to the cognate target during the course of the hybridization, and so target coverage at the end of the experiment is usually less than 5%. The target coverage using the methods of the invention are theoretically 100% because the label is in excess compared to the target analyte.

One of the advantages of 100% target coverage is that this allows the target analytes to be directly counted since each analyte is bound by a label. Using the methods of the invention, molecular species can be directly counted one by one. Direct counting, or a digital output, is preferable to the indirect quantitation methods used for microarrays because the data does not need to go through several intermediary transformations. If the detection equipment directly counts the number of particles emitted it is said to have a digital output, however if the direct counts undergo several intermediary transformations then the data has an analog output. Quantitation data from microarrays is known to be subject to several distortions due to data extrapolation.

Another consequence of low target coverage in microarrays is that highly sensitive equipment is needed to detect the low amount of signal. However, since

microarrays require a high dynamic range for signal detection, sensitivity is decreased because of the trade-off between sensitivity and dynamic range. The methods of the invention use a limited number of labels to create a large number of unique label combinations. This allow the detection method to operate in a narrow dynamic range. In contrast, microarray methods require a large dynamic range (4 orders of magnitude or more) to account for the large differences in abundance of the different molecular species. The low requirements in terms of dynamic range required by the methods of the invention will improve the sensitivity of the system since the trade-off between sensitivity and dynamic range is avoided.

The invention further provides a method of detecting a nucleic acid analyte. The method entails contacting a mixture of nucleic acid analytes under conditions sufficient for hybridization with a target specific probe having at least one genedigit where the genedigit has a set of three or more repeated sequences, then contacting that mixture with an anti-genedigit having a cognate set of at least two complimentary repeated sequences, and finally detecting a hybridized complex containing the analyte, target specific probe and anti-genedigit where the anti-genedigit hybridizes to the genedigit through a seqeunce having a complexity less than the number of hybridized base pairs.

The invention also provides a method of detecting a nucleic acid analyte as above where the anti-genedigit(s) each have a unique label.

The methods of the invention provide for detection of analytes in mixtures. The mixture can contain several types of analytes or the mixture can contain just one

type of analyte. In addition, the mixture could contain just a single copy of an analyte. If the target analyte has an unknown sequence or structure, a large population of target specific specifiers can be added to the mixture. This population can include specifiers with target specific regions of pre-determined sequence or structure or specifiers can be used with target specific regions of random sequence or structure. Alternatively, if the target analyte has a known sequence or structure, a particular specifier containing a region that will specifically bind to that sequence or structure can be used either alone or in combination with other specifiers.

The methods of the invention are suited to nucleic acid analytes as well as analytes with other structures. A population of specifiers can be generated for any analyte where a target specific region can be found that specifically interacts with that analyte. For example, a protein analyte could be bound specifically by a nucleic acid or a peptide or an antibody, all of which can be linked to nucleic acid genedigits. The target specific region can also be attached to amino acid containing genedigits. In addition, the ability of nucleic acid aptamers to bind a wide variety of analytes allows these structures to be used in the target specific regions of specifiers. Several combinations are possible as long as the specifiers specifically bind to the analytes.

The invention further provides a nucleic acid labeling kit, containing a set of genedigits, a set of anti-genedigits and a unique set of labels bound to a nucleic acid. The kit can include other reagents as well, for example, buffers for performing hybridization reactions, linkers, restriction endonucleases, and DNA

ligases. The kit also will include instructions for use of the labeling kit.

The labels and methods of the invention can be used for diagnostic and therapeutic purposes. Analytes, or combinations of analytes, that are diagnostic of a disease can be detected and quantified from a sample derived from a subject. Many different analytes can be analyzed at one time from a single sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample. In addition, the methods of the invention can provide information that determines a course of treatment for a patient. For example, the amount of a particular marker for a tumor can be accurately quantified from even a small sample from a patient. For some disease like breast cancer overexpression of certain genes, such as Her2-neu, indicate a more aggressive course of treatment will be needed.

It is understood that modifications which do not substantially affect the activity of the various embodiments of this invention are also included within the definition of the invention provided herein. Accordingly, the following examples are intended to illustrate but not limit the present invention.

EXAMPLE I

Generation of Unique Labels Using Two Different Labels

In this example, ten unique labels are made from two different fluorescent labels. First, ten unique templates of a 220-base pair single-stranded DNA are synthesized. The templates consist of a pre-determined

ratio of the following 20-base pair repeats:

5' (ACTCTCTCTCTCTCTCTCTC)n(GCTCTCTCTCTCTCTCTCTC)m 3' (SEQ ID NO:1) where n=1,2,3,4,5,6,7,8,9,10, m=1,2,3,4,5,6,7,8,9,10, and n+m=11.

The second strand is synthesized using the primer GAGAGAGAGA (SEQ ID NO:2), Klenow polymerase, DNA ligase, dGTP, dATP, dUTP-fluorescein and dCTP-rhodamine. After the reaction is complete the product is treated with S1 nuclease to digest the DNA with gaps, and the remaining full length DNA is then purified. The labeled nucleotides will be incorporated into the DNA in a unique ratio determined by the ratio of the two repeats. The end result is ten uniquely labeled nucleic acids where the set ratio of fluorescein to rhodamine is 1:10, 2:9, 3:8, 4:7, 5:6, 6:5, 7:4, 8:3, 9:2, and 10:1.

A linker oligonucleotide is ligated to the labeled DNA and then this linker is used to attach the labeled DNA to the branches of a dendrimer. The dendrimer has an oligonucleotide tag of 5 bases at the stem to facilitate binding of an anti-genedigit (see Example II) and tags of 10-base pairs at the branches to facilitate binding of labeled DNA.

EXAMPLE II

Generation of a Labeled Specifier

The specifiers are synthesized by ligating together one target specific sequence (synthetic oligonucleotide, peptide-nucleic acid (PNA), PCR product, or linked-nucleic acid (LNA)), and several "genedigits" (see Figure 1A). In this example, each specifier contains a unique combination of 4 different genedigits. This results in the generation of 10,000 possible unique specifiers.

The genedigits are synthetic oligonucleotides that contain only two of the natural bases, plus two bases that not found in nature: isocytidine and isoguanine. Such base composition ensures that the genedigits will not non-specifically hybridize with analytes in a complex mixture. The sequence of each genedigit is composed of 5 repeats of an 8-base pair core sequence (see Figure 1B). Each core sequence unit differs from the others by at least two bases.

In order to make 10,000 unique specifiers, forty different genedigits are synthesized and split into 4 groups containing 10 genedigits each. The genedigits of each group have a 5-base pair tag on each end. One genedigit from each group is present in each specifier. The genedigits are ligated with the help of adapter 10mer oligonucleotides that are complimentary to the 5-base pair tags that are common for each group. In this way, for a specifier with 4 genedigits, there will be 10x10x10x10 = 10,000 possible combinations. All specifiers also contain a biotin tag.

The genedigit serves as an attachment point for a label and so the number of labels synthesized corresponds to the number of genedigits. The genedigits are labeled through the use of anti-genedigits. An anti-genedigit sequence consisting of three 8-base pair repeats complimentary to the 8-base pair core repeat of the corresponding genedigit is ligated to the stem of a labeled dendrimer (from Example I).

The 24-base pair labeled anti-genedigit hybridizes to the 40-base pair genedigit sequence in the specifier in one of three different registers (see Figure 1C).

Thus when the label hybridized to the specifier it will do so through a 24-base pair sequence that only has a complexity of an 8-base pair sequence.

EXAMPLE III

Gene Expression Analysis Using Specifiers

In order to determine differences in gene expression between astrocytes and LPS-activated astrocytes, RNA is isolated from both populations of astrocytes using cell lysis in guanidine isothiocynine or phenol/chloroform. A population of specifiers is added to each RNA sample under conditions suitable for hybridization. The mRNA-specifier complexes are isolated with oligo dT beads and washed extensively to remove excess specifiers. The specifiers are eluted from the mRNA by digesting the mRNA with RNAse A. The specifiers are then are processed for labeling as described in Examples I and II and these labels are detected using a CCD camera. The number of specifiers corresponding to specific mRNAs from un-treated astrocytes is then compared to the specifier pattern from LPS-treated astrocytes. Since the sequence of the target specific region of the specifier is known, this identifies the genes that are differentially expressed between the two samples.

EXAMPLE IV

Microbial Detection Using Specifiers

The invention can be used to detect strains of microorganisms with known sequence in biological samples. Total DNA is extracted from a blood sample from a patient with a suspected microbial infection. The total DNA is then tagged with digoxygenin. The DNA is denatured and hybridized with a population of specifiers containing target regions specific for a particular microorganism,

or a panel of microorganisms, in a small volume (0.01 - 2.0 μl). The sample DNA-specifier complexes are isolated using anti-digoxygenin antibodies, and washed extensively to remove excess specifiers. The specifiers are then processed for labeling and imaging as described above.

EXAMPLE V

Label Configurations

This example shows a DNA probe molecule (specifier) that contains labels that are separated by a distance sufficient such that the labels can be individually resolved (Figure 2). For example, in the case of fluorescent labels, this large spatial separation of labels allows for spectral identification of each label individually.

As shown in Figure 2, a DNA probe (specifier) can be generated that contains four distinct fluorescent labels. These labels are separated by a distance of about one micron. In this example, the labels are separated by about 2,000 to 3,000 bases of DNA, however other polymers could be used to separate the labels as well.

A diverse population of unique labels can be generated by varying the fluorescent label at each of the four locations. For example, a set of 10 distinct fluorphores can be used to generate 10,000 unique labels. Specifically, for a probe with four different labels, there will be 10x10x10x10 = 10,000 possible combinations. Similarly, for a probe that has five different labels, there will be 100,000 possible combinations.

EXAMPLE VI

Flow-stretch Separation of Labels

This example shows a DNA probe molecule (specifier) with one label monomer which has been stretched on a coverslip using a flow stretch technique. The flow stretch technique can also be used for a probe molecule with multiple label monomers along the probe molecule.

A double stranded DNA probe was constructed with a single stranded section at a desired location for label attachment. In this example, the single stranded section is at the end of the probe molecule. The label molecule contains about 300 Cy3 fluorophores and a section of DNA that is complementary to the single stranded area of the probe molecule. After hybridization, the probe molecule is stained with YOYO1 dye to verify that the label is attached to the probe. The staining procedure is not necessary when multiple labels are attached. The probes with hybridized label were then stretched on a coverslip using a flow stretch technique (Henegariu et al., *supra)*. An image, figure 3, was obtained using an inverted fluorescent microscope. Probes can also be stretched using a receding meniscus technique (Yokota et al., *supra*).

EXAMPLE VII

Electro-stretch Separation of Labels

This example shows than an electric field can be used to align probe (specifier) molecules that are attached by one end to a surface. In this manner, labels that are along the probe molecule are spatially separated.

A flow cell was constructed that allowed imaging of the probe molecules through a coverslip. The flow cell also allowed for the exchange of solutions and the application of an electric field. In order to prevent bubbling, a gel was used to separate the electrodes from the solution. With gel separation of the electrodes and solution, higher voltages are possible which improves the alignment of the probe molecules. The coverslip was coated with bovine serum albumin (BSA) to minimize non-specific binding. DNA was added to the coverslip and attached to the coverslip primarily by the ends of the DNA. Un-bound DNA was washed away and the remaining bound DNA was stained with YOYO1. Imaging was performed using an epi-fluorescent microscope. When the voltage was off, tethered DNA can be seen to move randomly (Figure 4A). However, when the voltage is on, the DNA becomes aligned (Figure 4B).

EXAMPLE VIII

Cytometer Detection of Labels

Another way to align DNA probes (specifiers) spatially is by the use of a constriction in the flow of a liquid containing the DNA in conjunction with an oscillating electric field (Asbury, C.L., and van den Engh, G., Biophys. J. 74:1024-1030 (1998)). In this way, when the probe molecules pass through the constriction the labels are spatially separated (Figure 5). Confocal optics can allow for detection with enough spatial resolution to determine the sequence of the labels.

Throughout this application various publications have been referenced within parentheses. The disclosures of these publications in their entireties are hereby incorporated by reference in this application in order to

more fully describe the state of the art to which this invention pertains.

Although the invention has been described with reference to the disclosed embodiments, those skilled in the art will readily appreciate that the specific experiments detailed are only illustrative of the invention. It should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. A diverse population of labels, comprising about thirty or more unique labels, wherein each of said unique labels is bound to a nucleic acid molecule.

2. The diverse population of claim 1, wherein each of said unique labels further comprises about the same unit signal, or a multiple thereof.

3. The diverse population of claim 1, wherein each of said unique labels further comprises a mixture of two or more different labels.

4. The diverse population of claim 1, further comprising unique labels having a diversity selected from a group consisting of about 40, 60, 80, 100, 120 and 140.

5. The diverse population of claim 1, further comprising about 150.

6. The diverse population of claim 1, wherein said labels are fluorescent.

7. A diverse population of uniquely labeled probes, comprising about thirty or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid.

8. The diverse population of claim 7, wherein each of said unique labels further comprises about the same unit signal, or a multiple thereof.

9. The diverse population of claim 7, wherein each of said unique labels further comprises a mixture of two or more different labels.

10. The diverse population of claim 7, further comprising a diversity of different labels selected from a group consisting of about 50, 100, 200, 500, 1,000, 2,000, 5,000, $1x10^4$ and $3x10^4$.

11. The diverse population of claim 7, further comprising about $1x10^5$.

12. The diverse population of claim 7, wherein said labels are fluorescent.

13. The diverse population of claim 7, wherein said target specific nucleic acid further comprises said nucleic acid bound to said unique label.

14. The diverse population of claim 7, further comprising two attached populations of nucleic acids, one population of nucleic acids comprising said thirty or more target specific nucleic acid probes, and a second population of nucleic acids comprising said nucleic acid bound by a unique label.

15. The diverse population of claim 7, further comprising a nucleic acid species selected from the group consisting of a specifier, an antispecifier, a genedigit, an anti-genedigit and a dendrimer.

16. A method of producing a population of labels, comprising synthesizing a population of nucleic acids, each of said nucleic acids within said synthesized population having bound a predetermined ratio of at least two different labels, wherein said population of labels has a diversity of predetermined ratios of about 30 or more.

17. The method of claim 16, wherein each of said nucleic acids bound by a predetermined ratio of at least two labels further comprises about the same unit signal.

18. The method of claim 16, wherein said diversity is selected from the group consisting of 40, 60, 80, 100, 120, 140, and 150.

19. The method of claim 16, wherein said diversity is selected from the group consisting of 200, 500, 2,000, 5,000, $1x10^4$, $3x10^4$ and $1x10^5$.

20. The method of claim 16, wherein said labels are fluorescent.

21. The method of claim 16, wherein each of said nucleic acids within said synthesized population further comprise an anti-genedigit.

22. The method of claim 16, wherein each of said nucleic acids within said synthesized population further comprise a target specific probe.

23. A method of attaching a label to a nucleic acid probe, comprising hybridizing a nucleic acid probe having a genedigit to an anti-genedigit having a label, said genedigit comprising a set of three or more repeat sequences, said anti-genedigit comprising a cognate set of at least two complimentary repeat sequences, wherein said anti-genedigit specifically hybridizes to said genedigit through a sequence having a complexity less than the number of hybridized base pairs.

24. The method of claim 23, wherein said nucleic acid probe further comprises a specifier.

25. The method of claim 24, wherein said specifier further comprises four or more genedigits.

26. The method of claim 24, wherein said specifier further comprises five or more genedigits.

27. The method of claim 23, wherein said genedigit further comprises a set of four or more repeat sequences.

28. The method of claim 23, wherein said anti-genedigit further comprises a cognate set of three or more complimentary repeat sequences.

29. The method of claim 23, wherein said repeat sequences or complimentary repeat sequences further comprise about 8 nucleotides.

30. The method of claim 23, wherein said complexity is about 8.

31. The method of claim 23, wherein said number of hybridized base pairs is about 24.

32. The method of claim 23, further comprising hybridizing a population of nucleic acid probes having different genedigits to a population of different anti-genedigits each having unique labels to produce a population of uniquely labeled nucleic acid probes.

33. The method of claim 32, wherein said nucleic acid probe further comprises a specifier.

34. A method of producing a population of nucleic acid probes, comprising:

(a) producing a first population of nucleic acids comprising two or more target specific probes each having at least one genedigit, said genedigit having a set of three or more repeated sequences;

(b) producing a second population of nucleic acids comprising an anti-genedigit having a cognate set of at least two complimentary repeated sequences, and

(c) hybridizing said first and second populations of nucleic acids to produce a population of target specific probes attached to an anti-genedigit, wherein said anti-genedigit hybridizes to said genedigit through a sequence having a complexity less than the number of hybridized base pairs.

35. The method of claim 34, wherein said target specific probes further comprise a specifier.

36. The method of claim 35, wherein said specifier further comprises four or more genedigits.

37. The method of claim 35, wherein said specifier further comprises five or more genedigits.

38. The method of claim 34, wherein said genedigit further comprises a set of four or more repeat sequences.

39. The method of claim 34, wherein said anti-genedigit further comprises a cognate set of three complimentary repeat sequences.

40. The method of claim 34, wherein said repeat sequences or complimentary repeat sequences further comprise about 8 nucleotides.

41. The method of claim 34, wherein said complexity is about 8.

42. The method of claim 34, wherein said number of hybridized base pairs is about 24.

43. The method of claim 34, wherein said two or more target specific probes further comprise at least one different genedigit.

44. The method of claim 34, wherein said population of anti-genedigits further comprises two or more different anti-genedigits.

45. The method of claim 34, wherein said anti-genedigit further comprises a label.

46. The method of claim 44, wherein said two or more different anti-genedigits further comprise unique labels.

47. The method of claim 34, further comprising hybridizing a population of target specific probes having different genedigits to a population of different anti-genedigits each having a unique label to produce a population of uniquely labeled nucleic acid probes.

48. The method of claim 47, wherein said target specific nucleic acid probe further comprises a specifier.

49. A method of producing a population of uniquely labeled nucleic acid probes, comprising:

(a) synthesizing a population of target specific nucleic acid probes each having a different specifier;

(b) synthesizing a corresponding population of anti-genedigits each having a unique label, said population having a diversity sufficient to uniquely hybridize to genedigits within said specifiers, and

(c) hybridizing said populations of target nucleic acid probes to said anti-genedigits, to produce a population wherein each of said target specific probes is uniquely labeled.

50. A method of detecting a nucleic acid analyte, comprising:

(a) contacting a mixture of nucleic acid analytes with a plurality of target specific probes each attached to a unique label bound to a nucleic acid under conditions sufficient for hybridization of said probes to said target, and

(b) measuring a signal from one or more said target specific probes hybridized to an analyte, wherein said signal uniquely identifies the analyte species.

51. The method of claim 50, further comprising stretching of said target specific probes hybridized to an analyte using a flow-stretch, receding meniscus, or electro-stretch technique, thereby spatially separating labels within a unique label.

52. The method of claim 50, wherein each of said unique labels further comprises about the same unit signal, or multiple thereof.

53. The method of claim 50, wherein each of said unique labels further comprises a mixture of two or more different labels.

54. The method of claim 50, wherein said plurality of target specific probes further comprises different target specific probes selected from the group consisting of about 50, 100, 200, 500, 1,000, 2,000, 5,000, $1x10^4$, $3x10^4$ and $1x10^5$.

55. The method of claim 50, wherein said plurality of target specific probes further comprises at least one target specific probe for each nucleic acid analyte in said mixture.

56. The method of claim 50, wherein said mixture further comprises an expressed RNA population or DNA copy thereof.

57. The method of claim 50, wherein said labels are fluorescent.

58. The method of claim 50, wherein said target specific nucleic acid further comprises said nucleic acid bound to said unique label.

59. The diverse population of claim 50, further comprising two attached populations of nucleic acids, one population of nucleic acids comprising said plurality of target specific nucleic acid probes, and a second population of nucleic acids comprising said nucleic acid bound by a unique label.

60. The diverse population of claim 50, further comprising a nucleic acid species selected from the group consisting of a specifier, an antispecifier, a genedigit, an anti-genedigit and a dendrimer.

61. The method of claim 50, further comprising detecting a single copy of an analyte within said mixture.

62. A method of detecting a nucleic acid analyte, comprising:

(a) contacting a mixture of nucleic acid analytes under conditions sufficient for hybridization with a target specific probe having at least one genedigit, said genedigit having a set of three or more repeated sequences;

(b) contacting said mixture under conditions sufficient for hybridization with an anti-genedigit having a cognate set of at least two complimentary repeated sequences, and

(c) detecting a hybridized complex comprising said analyte, target specific probe and said anti-genedigit, wherein said anti-genedigit hybridizes to said genedigit through a sequence having a complexity less than the number of hybridized base pairs.

63. The method of claim 62, wherein detecting said hybridized complex comprising said analyte, target specific probe and said anti-genedigit comprises the use of a flow-stretch, receding meniscus, or electro-stretch technique.

64. The method of claim 62, wherein said target specific probe further comprises a specifier.

65. The method of claim 64, wherein said specifier further comprises four or more genedigits.

66. The method of claim 64, wherein said specifier further comprises five or more genedigits.

67. The method of claim 62, wherein said genedigit further comprises a set of four or more repeat sequences.

68. The method of claim 62, wherein said anti-genedigit further comprises a cognate set of three complimentary repeat sequences.

69. The method of claim 62, wherein said repeat sequences or complimentary repeat sequences further comprise about 8 nucleotides.

70. The method of claim 62, wherein said complexity is about 8.

71. The method of claim 62, wherein said number of hybridized base pairs is about 24.

72. The method of claim 62, further comprising two or more target specific probes

73. The method of claim 72, further comprising at least one different genedigit.

74. The method of claim 72, further comprising two or more different anti-genedigits.

75. The method of claim 62, wherein said anti-genedigit further comprises a label.

76. The method of claim 74, wherein said two or more different anti-genedigits further comprise unique labels.

77. The method of claim 62, further comprising a plurality of target specific probes having different genedigits and a plurality of different anti-genedigits each having a unique label.

78. The method of claim 77, wherein said target specific probes further comprise a specifier.

79. A method of detecting a nucleic acid analyte, comprising:

(a) contacting a mixture of nucleic acid analytes under conditions sufficient for hybridization with a plurality of target specific nucleic acid probes each having a different specifier;

(b) contacting said mixture under conditions sufficient for hybridization with a corresponding plurality of anti-genedigits each having a unique label, said plurality of anti-genedigits having a diversity sufficient to uniquely hybridize to genedigits within said specifiers, and

(c) uniquely detecting a hybridized complex between one or more analytes in said mixture, a target specific probe, and an anti-genedigit.

80. The method of claim 79, wherein said uniquely detecting a hybridized complex between one or more analytes in said mixture, a target specific probe, and an anti-genedigit comprises the use of a flow-stretch, receding meniscus, or electro-stretch technique.

81. A nucleic acid labeling kit, comprising a set of genedigits, a set of anti-genedigits and a unique set of labels bound to a nucleic acid.

82. The kit of claim 81, further comprising a specifier.

83. The kit of claim 81, further comprising a dendrimer.

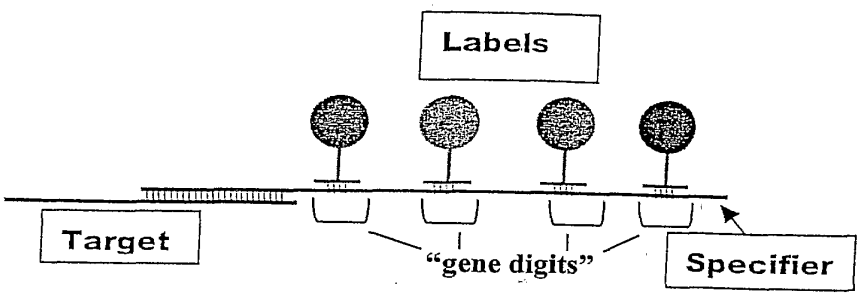
84. The kit of claim 81, wherein said unique label further comprises a ratio of two different labels.

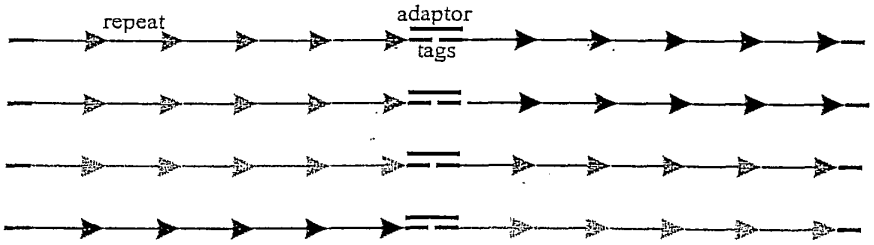
85. The kit of claim 81, wherein said unique label is fluorescent.

86. The kit of claim 81, further comprising a diverse population of unique labels.

87. The kit of claim 86, wherein said diversity is selected from the group consisting of 50, 100, 150, 200, 500, 1,000, 2,000, 5,000, $1x10^4$, $3x10^4$ and $1x10^5$.

Figure 1


Labels
Target
"gene digits"
Specifier


repeat
adaptor
tags

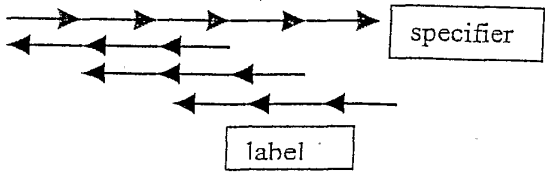

specifier
label

Figure 2

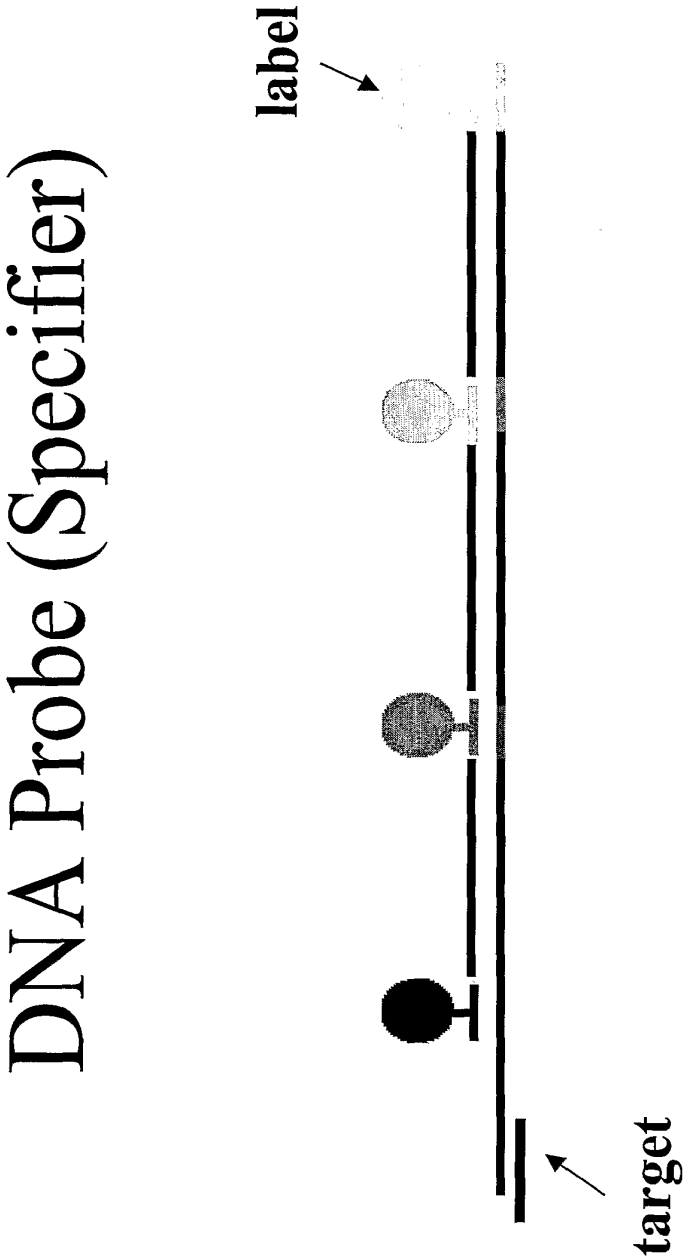


Figure 3

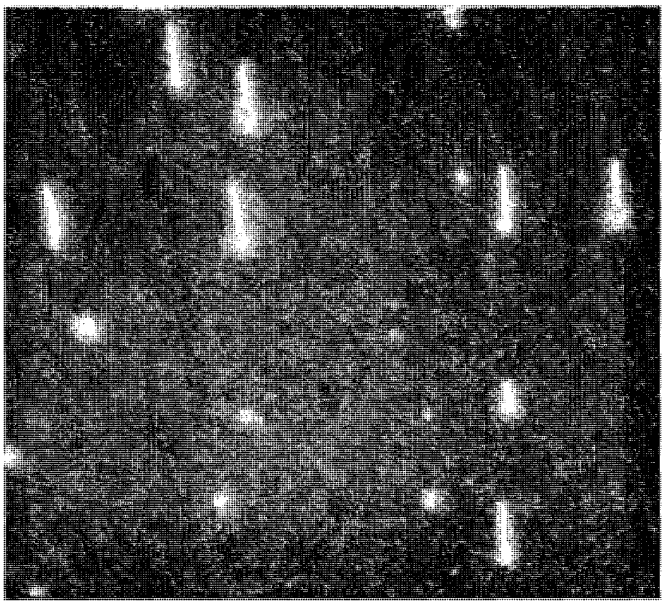
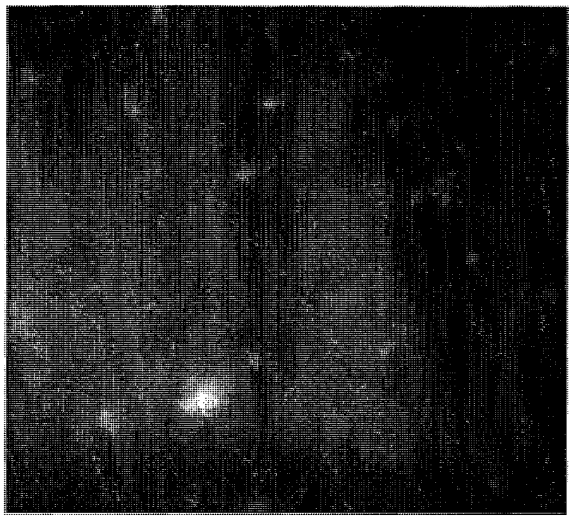


Figure 4

A



B

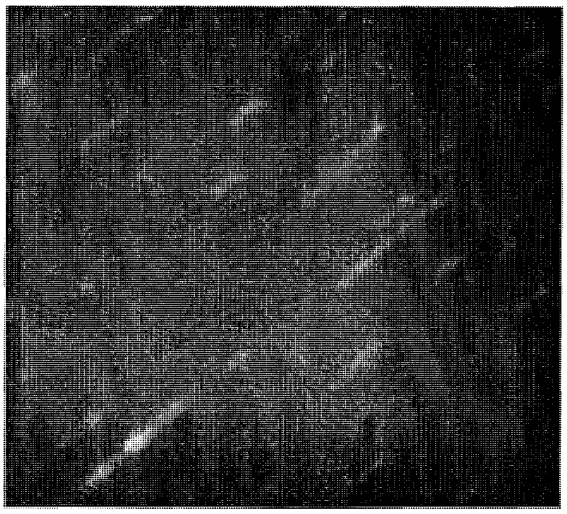
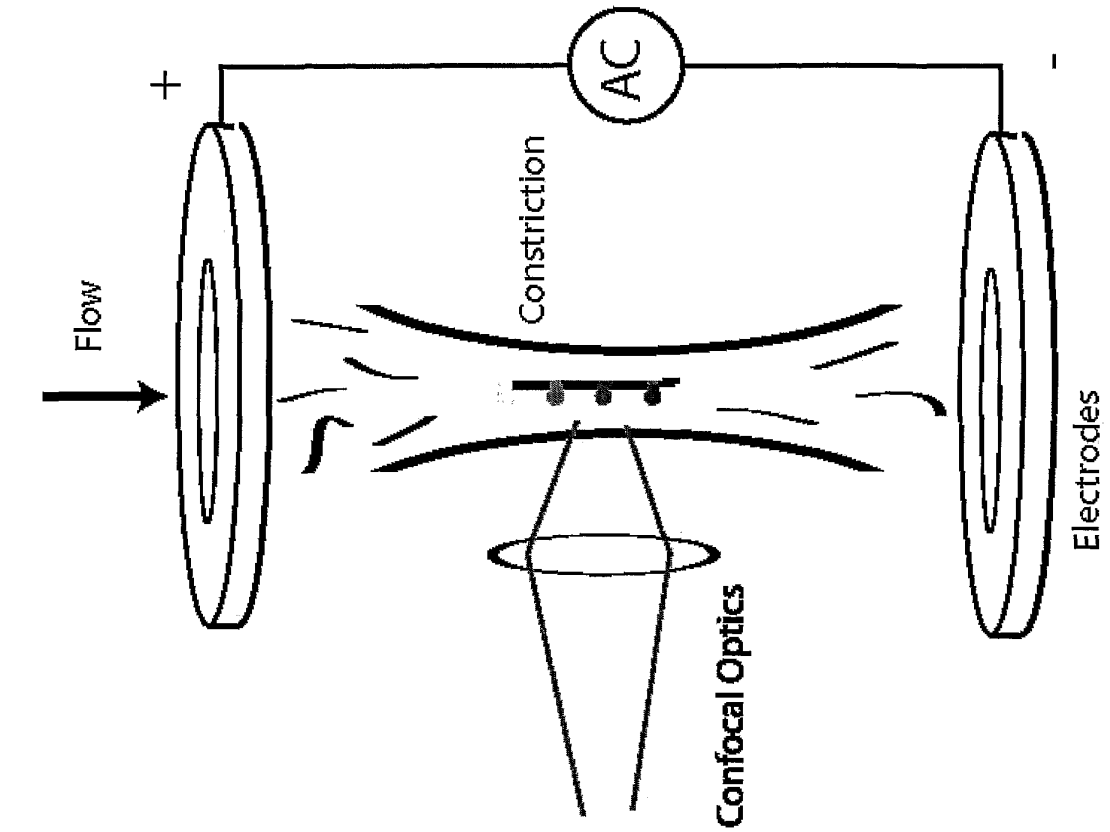


Figure 5



DNA alignment in AC field

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 January 2003 (16.01.2003)

PCT

(10) International Publication Number
WO 03/003810 A3

(51) International Patent Classification[7]: C07H 21/02, C12Q 1/68

(21) International Application Number: PCT/US02/21278

(22) International Filing Date: 3 July 2002 (03.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/898,743 3 July 2001 (03.07.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/898,743 (CIP)
Filed on 3 July 2001 (03.07.2001)

(71) Applicant *(for all designated States except US)*: **THE INSTITUTE FOR SYSTEMS BIOLOGY** [US/US]; 1441 North 34th Street, Seattle, WA 98103-8904 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **DIMITROV, Krassen** [BG/US]; 10759 Burke Avenue N., Seattle, WA 98133 (US). **DUNAWAY, Dwayne** [US/US]; 6548 39th Avenue N.E., Seattle, WA 98115 (US).

(74) Agents: **CADENA, Deborah, L.** et al.; Campbell & Flores LLP, 7th floor, 4370 La Jolla Village Drive, San Diego, CA 92122 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
27 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/003810 A3

(54) Title: METHODS FOR DETECTION AND QUANTIFICATION OF ANALYTES IN COMPLEX MIXTURES

(57) Abstract: The invention provides a diverse population of uniquely labeled probes, containing about thirty or more target specific nucleic acid probes each attached to a unique label bound to a nucleic acid. Also provided is a method of producing a population of uniquely labeled nucleic acid probes. The method consists of (a) synthesizing a population of target specific nucleic acid probes each having a different specifier; (b) synthesizing a corresponding population of anti-genedigits each having a unique label, the population having a diversity sufficient to uniquely hybridize to genedigits within the specifiers, and (c) hybridizing the populations of target nucleic acid probes to the anti-genedigits, to produce a population in which each of the target specific probes is uniquely labeled. Also provided is a method of detecting a nucleic acid analyte. The method consists of (a) contacting a mixture of nucleic acid analytes under conditions sufficient for hybridization with a plurality of target specific nucleic acid probes each having a different specifier; (b) contacting the mixture under conditions sufficient for hybridization with a corresponding plurality of anti-genedigits each having a unique label, the plurality of anti-genedigits having a diversity sufficient to uniquely hybridize to genedigits within the specifiers, and (c) uniquely detecting a hybridized complex between one or more analytes in the mixture, a target specific probe, and an anti-genedigit.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/21278

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : CO7H 21/02; C12Q 1/68
US CL : 536/25.32; 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/25.32; 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 99/19515 A1 (CHANDLER et al), 22 April 1999, entire document, especially page 29, lines 24-29, page 30, lines 1-12, page 31, lines 19-29.	1-13, 16-20, --- 14-15, 21-87
Y	US 5,487,973 A (NILSEN et al) 30 January 1996, entire document	14-15, 21-87
Y	US 5,401,847 A (GLAZER et al) 28 March 1995, entire document.	21-87

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 November 2002 (04.11.2002)	27 JAN 2003
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Suryaprabha Chunduru Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US02/21278

Continuation of B. FIELDS SEARCHED Item 3:
Biosis, Embase, Medline, Caplus, Lifesci databases
Search terms: multiple labels or markers, hybridization, probes or oligos

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2008 (12.09.2008)

PCT

(10) International Publication Number
WO 2008/108843 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2007/063103

(22) International Filing Date: 1 March 2007 (01.03.2007)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **GEN-PROBE INCORPORATED** [US/US]; Mail Stop 1 / Patent Department, 10210 Genetic Center Drive, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BECKER, Michael, M.** [US/US]; 12819 Meadowdale Lane, San Diego, CA 92131 (US). **LAM, Wai-Chung** [US/US]; 2808 State Street, #a, San Diego, CA 92013 (US). **LIVEZEY, Kristin, W.** [US/US]; 551-A Hygeia Avenue, Encinitas, CA 92024 (US). **BRENTANO, Steven, T.** [US/US]; 8426 Mesa Ridge Road, Santee, CA 92071 (US). **KOLK, Daniel, P.** [US/US]; 2118 Prestige Road, Ramona, CA 92065 (US). **SCHRODER, Astrid, R.W.** [DE/US]; 11203 Carmel Creek Road, #5, San Diego, CA 92130 (US).

(74) Agent: **CAPPELLARI, Charles**; Gen-Probe Incorporated, Mail Stop 1 / Patent Department, 10210 Genetic Center Drive, San Diego, CA 92121 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: METHODS AND KITS FOR AMPLIFYING DNA

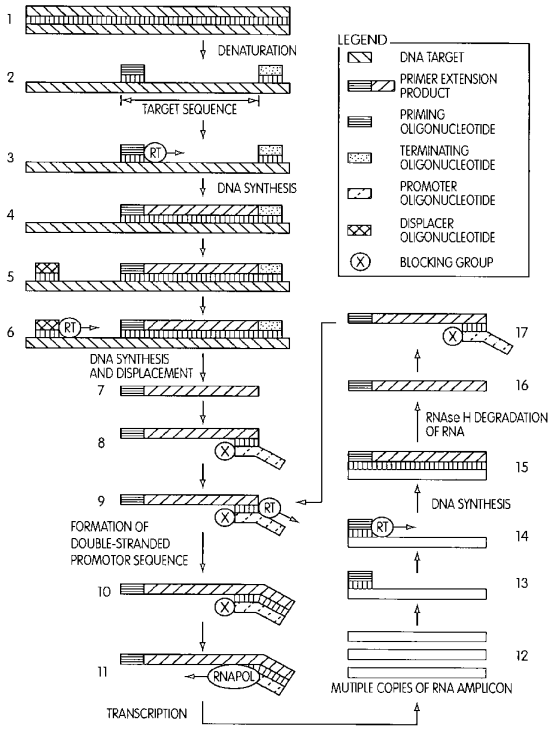


FIG. 1D

(57) Abstract: Novel methods of synthesizing multiple copies of a target nucleic acid sequence which are autocatalytic are disclosed (/. e., able to cycle automatically without the need to modify reaction conditions such as temperature, pH, or ionic strength and using the product of one cycle in the next one). In particular, methods of nucleic acid amplification are disclosed which are robust and efficient, while reducing the appearance of side-products. In general, the methods use priming oligonucleotides that target only one sense of a target nucleic acid, a promoter oligonucleotide modified to prevent polymerase extension from its 3 '-terminus and, optionally, a means for terminating a primer extension reaction, to amplify RNA or DNA molecules in vitro, while reducing or substantially eliminating the formation of side-products. The disclosed methods minimizes or substantially eliminate the emergence of side-products, thus providing ahigh level of specificity. Furthermore, the appearance of side-products can complicate the analysis of the amplification reaction by various molecular detection techniques. The disclosed methods minimize or substantially eliminate this problem, thus providing enhanced levels of sensitivity.

WO 2008/108843 A2

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

METHODS AND KITS FOR AMPLIFYING DNA

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application Serial No. 11/213,519, filed August 26, 2005, now pending, which claims the benefit of U.S. Provisional Application No. 60/604,830, filed August 27, 2004, and U.S. Provisional Application No. 60/639,110, filed December 23, 2004, the contents of each of which applications are hereby incorporated by reference herein.

FIELD OF THE INVENTION

This invention relates to methods, reaction mixtures, compositions and kits for producing multiple copies of a specific nucleic acid sequence or "target sequence" which may be present either alone or as a component, large or small, of a homogeneous or heterogeneous mixture of nucleic acids. The mixture of nucleic acids may be that found in a sample taken for diagnostic testing, for screening of blood products, for food, water, industrial or environmental testing, for research studies, for the preparation of reagents or materials for other processes such as cloning, or for other purposes.

The selective amplification of specific nucleic acid sequences is of value in increasing the sensitivity of diagnostic and other detection assays while maintaining specificity; increasing the sensitivity, convenience, accuracy and reliability of a variety of research procedures; and providing ample supplies of specific oligonucleotides for various purposes.

BACKGROUND OF THE INVENTION

The detection and/or quantitation of specific nucleic acid sequences is an important technique for identifying and classifying microorganisms, diagnosing infectious diseases, detecting and characterizing genetic abnormalities, identifying genetic changes associated with cancer, studying genetic susceptibility to disease, and measuring response to various types of treatment. Such procedures are also useful in detecting and quantitating microorganisms in foodstuffs, water, industrial and environmental samples, seed stocks, and other types of material where the presence of specific microorganisms may need to be monitored. Other applications are found in the forensic sciences, anthropology, archaeology, and biology where measurement of the relatedness of nucleic acid sequences has been used to identify criminal suspects, resolve paternity disputes, construct genealogical and phylogenetic trees, and aid in classifying a variety of life forms.

A number of methods to detect and/or quantitate nucleic acid sequences are well known in the art. These include hybridization to a labeled probe, and various permutations

of the polymerase chain reaction (PCR), coupled with hybridization to a labeled probe. *See, e.g.*, Mullis *et al.*, "Process for Amplifying, Detecting and/or Cloning Nucleic Acid Sequences," U.S. Patent No. 4,683,195; Mullis, "Process for Amplifying Nucleic Acid Sequences," U.S. Patent No. 4,683,202; Mullis *et al.*, "Process for Amplifying, Detecting and/or Cloning Nucleic Acid Sequences," U.S. Patent No. 4,800,159; Mullis *et al.* (1987) *Meth. Enzymol.* 155, 335-350; and Murakawa *et al.* (1988) *DNA* 7, 287-295. The requirement of repeated cycling of reaction temperature between several different and extreme temperatures is a disadvantage of the PCR procedure. In order to make PCR convenient, expensive programmable thermal cycling instruments are required.

Additionally, Transcription-Mediated Amplification (TMA) methods may be used to synthesize multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autocatalytically generate additional copies. *See, e.g.*, Kacian *et al.*, "Nucleic Acid Sequence Amplification Methods," U.S. Patent No. 5,399,491, and Kacian *et al*, "Nucleic Acid Sequence Amplification Methods," U.S. Patent No. 5,824,518, the contents of each of which patents are hereby incorporated by reference herein. TMA is useful for generating copies of a nucleic acid target sequence for purposes which include assays to quantitate specific nucleic acid sequences in clinical, environmental, forensic and similar samples, cloning and generating probes. TMA is a robust and highly sensitive amplification system with demonstrated efficacy. TMA overcomes many of the problems associated with PCR-based amplification systems. In particular, temperature cycling is not required. Other transcription-based amplification methods are disclosed by Malek *et al.*, "Enhanced Nucleic Acid Amplification Process," U.S. Patent No. 5,130,238; Davey *et al.*, "Nucleic Acid Amplification Process," U.S. Patent No. 5,409,818; Davey *et al.*, "Method for the Synthesis of Ribonucleic Acid (RNA)," U.S. Patent No. 5,466,586; Davey *et al.*, "Nucleic Acid Amplification Process," U.S. Patent No. 5,554,517; Burg *et al.*, "Selective Amplification of Target Polynucleotide Sequences," U.S. Patent No. 6,090,591; and Burg *et al.*, "Selective Amplification of Target Polynucleotide Sequences," U.S. Patent No. 6,410,276.

An inherent result of highly sensitive nucleic amplification systems is the emergence of side-products. Side-products include molecules which may, in some systems, interfere with the amplification reaction, thereby lowering specificity. This is because limited amplification resources, including primers and enzymes needed in the formation of primer extension and transcription products are diverted to the formation of side-products. In some situations, the appearance of side-products can also complicate the analysis of amplicon production by various molecular techniques.

Accordingly, there remains a need in the art for a robust nucleic acid amplification system to synthesize multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH which reduces the appearance of side-products, thereby increasing specificity and improving detection and quantitation of amplification products.

SUMMARY OF THE INVENTION

The present invention is directed to novel methods of synthesizing multiple copies of a target sequence which are autocatalytic (*i.e.*, able to cycle automatically without the need to modify reaction conditions such as temperature, pH, or ionic strength and using the product of one cycle in the next one). In particular, the present invention discloses methods of nucleic acid amplification which are robust and efficient, while reducing the appearance of side-products. The methods use a priming oligonucleotide, a promoter oligonucleotide modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, includes a 3'-blocking moiety) and, optionally, a displacer oligonucleotide, a binding molecule and/or a 3'-blocked extender oligonucleotide, to amplify RNA or DNA molecules *in vitro*. Primers used in the disclosed methods target only one sense of a target nucleic acid. The methods of the present invention minimize or substantially eliminate the emergence of side-products, thus providing a high level of specificity. Furthermore, the appearance of side-products can complicate the analysis of the amplification reaction by various molecular detection techniques. The present invention minimizes or substantially eliminates this problem, thus providing an enhanced level of sensitivity.

In one embodiment, the present invention is drawn to a method of synthesizing multiple copies of a target sequence comprising treating a target nucleic acid which comprises an RNA target sequence with a priming oligonucleotide and a binding molecule (*e.g.*, terminating oligonucleotide or digestion oligonucleotide), where the priming oligonucleotide hybridizes to the 3'-end of the target sequence such that a primer extension reaction can be initiated therefrom, and where the binding molecule binds to the target nucleic acid adjacent to or near the 5'-end of the target sequence (by "adjacent to" is meant that the binding molecule binds to a base of the target nucleic acid next to the 5'-terminal base of the target sequence and fully 5' to the target sequence); extending the priming oligonucleotide in a primer extension reaction with a DNA polymerase, *e.g.*, reverse transcriptase, to give a DNA primer extension product complementary to the target sequence, where the primer extension product has a 3'-end which is determined by the binding molecule, where the 3'-end of the primer extension product is complementary to the 5'-end of the target sequence; separating the primer extension product from the target sequence using an enzyme which selectively

degrades the target sequence, *e.g.*, an enzyme with an RNAse H activity; treating the primer extension product with a promoter oligonucleotide comprising first and second regions, where the first region hybridizes to a 3'-region of the primer extension product to form a promoter oligonucleotide:primer extension product hybrid, where the second region comprises a promoter for an RNA polymerase and is situated 5' to the first region, and where the promoter oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, a blocking moiety is situated at the 3'-terminus of the promoter oligonucleotide which prevents polymerase extension); extending the 3'-end of the primer extension product in the promoter oligonucleotide:primer extension product hybrid to add a sequence complementary to the second region of the promoter oligonucleotide; and transcribing from the promoter oligonucleotide:primer extension product hybrid multiple RNA products complementary to the primer extension product using an RNA polymerase which recognizes the promoter in the promoter oligonucleotide and initiates transcription therefrom. According to this embodiment, the base sequences of the resulting RNA products are substantially identical to the base sequence of the target sequence. In a preferred aspect of this embodiment, the activity of the DNA polymerase is substantially limited to the formation of primer extension products comprising the priming oligonucleotide. In yet another preferred aspect of this embodiment, the formation of side-products in the method is substantially less than if said promoter oligonucleotide was not modified to prevent the initiation of DNA synthesis therefrom. According to yet another preferred aspect of this embodiment, if an oligonucleotide used in the amplification reaction comprises a promoter for an RNA polymerase, then that oligonucleotide further comprises a blocking moiety situated at its 3'-terminus to prevent the initiation of DNA synthesis therefrom.

A second embodiment of the present invention is drawn to a method of synthesizing multiple copies of a target sequence, where the method comprises treating a target nucleic acid comprising an RNA target sequence with a priming oligonucleotide which hybridizes to the 3'-end of the target sequence such that a primer extension reaction can be initiated therefrom; extending the priming oligonucleotide in a primer extension reaction with a DNA polymerase, *e.g.*, reverse transcriptase, to give a first DNA primer extension product having an indeterminate 3'-end and comprising a base region complementary to the target sequence; separating the first primer extension product from the target nucleic acid using an enzyme which selectively degrades that portion of the target nucleic acid which is complementary to the first primer extension reaction, *e.g.*, an enzyme with an RNAse H activity; treating the first primer extension product with a promoter oligonucleotide comprising first and second regions, where the first region hybridizes to a 3'-region of the first primer extension product to form a promoter oligonucleotide:first primer extension product hybrid, where the second

region comprises a promoter for an RNA polymerase and is situated 5' to the first region, and where the promoter oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, a blocking moiety is situated at the 3'-terminus of the promoter oligonucleotide which prevents polymerase extension); and transcribing from the promoter oligonucleotide:first primer extension product hybrid multiple first RNA products complementary to at least a portion of the first primer extension product using an RNA polymerase which recognizes the promoter and initiates transcription therefrom, where the base sequences of the resulting first RNA products are substantially identical to the base sequence of the target sequence. In a preferred aspect of this embodiment, the activity of the DNA polymerase in the method is substantially limited to the formation of primer extension products comprising the priming oligonucleotide. In yet another preferred aspect of this embodiment, the formation of side-products in the method is substantially less than if the promoter oligonucleotide was not modified to prevent the initiation of DNA synthesis therefrom. According to yet another preferred aspect of this embodiment, if an oligonucleotide used in the amplification reaction comprises a promoter for an RNA polymerase, then that oligonucleotide further comprises a blocking moiety situated at its 3'-terminus to prevent the initiation of DNA synthesis therefrom.

This embodiment is preferably drawn to the further steps of treating a first RNA product transcribed from the promoter oligonucleotide:first primer extension product with the priming oligonucleotide described above to form a priming oligonucleotide:first RNA product hybrid such that a primer extension reaction can be initiated from the priming oligonucleotide; extending the priming oligonucleotide in a primer extension reaction with a DNA polymerase, *e.g.*, reverse transcriptase, to give a second DNA primer extension product complementary to the first RNA product, where the second primer extension product has a 3'-end which is complementary to the 5'-end of the first RNA product; separating the second primer extension product from the first RNA product using an enzyme which selectively degrades the first RNA product, *e.g.*, an enzyme with an RNAse H activity; treating the second primer extension product with the promoter oligonucleotide described above to form a promoter oligonucleotide:second primer extension product hybrid; extending the 3'-end of the second primer extension product in the promoter oligonucleotide:second primer extension product hybrid to add a sequence complementary to the second region of the promoter oligonucleotide; and transcribing from the promoter oligonucleotide:second primer extension product hybrid multiple second RNA products complementary to the second primer extension product using an RNA polymerase, where the base sequences of the second RNA products are substantially identical to the base sequence of the target sequence.

A third embodiment of the present invention is drawn to a method of synthesizing multiple copies of a target sequence comprising treating a target nucleic acid comprising a DNA target sequence with a promoter oligonucleotide comprising first and second regions, where the first region hybridizes to the 3'-end of the target sequence to form a promoter oligonucleotide:target nucleic acid hybrid, where the second region comprises a promoter for an RNA polymerase and is situated 5' to the first region, and where the promoter oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, a blocking moiety is situated at the 3'-terminus of the promoter oligonucleotide); transcribing from the promoter oligonucleotide:target nucleic acid hybrid multiple first RNA products comprising a base region complementary to the target sequence using an RNA polymerase which recognizes the promoter and initiates transcription therefrom; treating the first RNA products with a priming oligonucleotide which hybridizes to a 3'-region of the first RNA products such that a primer extension reaction may be initiated therefrom; extending the priming oligonucleotide in the primer extension reaction with a DNA polymerase, *e.g.*, reverse transcriptase, to give a DNA primer extension product complementary to at least a portion of the first RNA products, where the primer extension product has a 3'-end which is complementary to the 5'-end of the first RNA products; separating the primer extension product from the first RNA product using an enzyme which selectively degrades the first RNA product (*e.g.*, an enzyme with an RNAse H activity); treating the primer extension product with the promoter oligonucleotide described above to form a promoter oligonucleotide:primer extension product hybrid; and transcribing from the promoter oligonucleotide:primer extension product hybrid multiple second RNA products complementary to the primer extension product using an RNA polymerase, wherein the base sequences of the second RNA products are substantially complementary to the base sequence of the target sequence. In a preferred aspect of this embodiment, the activity of the DNA polymerase in the method is substantially limited to the formation of primer extension products comprising the priming oligonucleotide. In yet another preferred aspect of this embodiment, the formation of side-products in the method is substantially less than if the promoter oligonucleotide was not modified to prevent the initiation of DNA synthesis therefrom. According to yet another preferred aspect of this embodiment, if an oligonucleotide used in the amplification reaction comprises a promoter for an RNA polymerase, then that oligonucleotide further comprises a blocking moiety situated at its 3'-terminus to prevent the initiation of DNA synthesis therefrom. Furthermore, any method of this embodiment may include extending the 3'-end of the primer extension product in the promoter oligonucleotide:primer extension product hybrid described above to add a sequence complementary to the second region of the promoter oligonucleotide.

A fourth embodiment of the present invention is drawn to a method of synthesizing multiple copies of a target sequence comprising treating a target nucleic acid which comprises a DNA target sequence with a priming oligonucleotide, where the priming oligonucleotide hybridizes to the 3'-end of the target sequence such that a primer extension reaction can be initiated therefrom; extending the priming oligonucleotide in a primer extension reaction with a DNA polymerase, *e.g.*, reverse transcriptase, to give a first DNA primer extension product, where at least a portion of the first primer extension product is complementary to the target sequence; treating the primer extension product with a promoter oligonucleotide comprising first and second regions, where the first region comprises a base sequence which corresponds to a region at the 5'-end of the target sequence and which hybridizes to the first primer extension product to form a promoter oligonucleotide:first primer extension product hybrid, where the second region comprises a promoter for an RNA polymerase and is situated 5' to the first region, and where the promoter oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, a blocking moiety is situated at the 3'-terminus of the promoter oligonucleotide which prevents polymerase extension); and transcribing from the promoter oligonucleotide:first primer extension product hybrid multiple first RNA products complementary to at least a portion of the first primer extension product using an RNA polymerase which recognizes the promoter in the promoter oligonucleotide and initiates transcription therefrom. Provided that if the first primer extension product has a defined 3'-end, then the method further comprises treating the target nucleic acid with a binding molecule which binds to the target nucleic acid adjacent to or near the 5'-end of the target sequence. Further provided that the priming oligonucleotide does not include an RNA region which hybridizes to the target nucleic acid and which is selectively degraded by an enzyme activity when hybridized to the target nucleic acid. According to this embodiment, the base sequences of the resulting first RNA products are substantially identical to the base sequence of the target sequence. In a preferred aspect of this embodiment, the target nucleic acid is part of a double-stranded complex that is exposed to conditions sufficient to denature the complex (*e.g.*, heat and/or chemical denaturants) prior to extending the priming oligonucleotide in a primer extension reaction. In another preferred aspect of this embodiment, the activity of the DNA polymerase is substantially limited to the formation of primer extension products comprising the priming oligonucleotide. In yet another preferred aspect of this embodiment, the formation of side-products in the method is substantially less than if said promoter oligonucleotide was not modified to prevent the initiation of DNA synthesis therefrom. In

yet a further preferred aspect of this embodiment, if an oligonucleotide used in the amplification reaction comprises a promoter for an RNA polymerase, then that oligonucleotide further comprises a blocking moiety situated at its 3'-terminus to prevent the initiation of DNA synthesis therefrom.

The method of this embodiment is preferably drawn to the further steps of treating the first RNA products transcribed from the promoter oligonucleotide:first DNA primer extension product with the priming oligonucleotide described above to form a priming oligonucleotide:first RNA product hybrid such that a primer extension reaction can be initiated from the priming oligonucleotide; extending the priming oligonucleotide in a primer extension reaction with a DNA polymerase, *e.g.*, reverse transcriptase, to give a second DNA primer extension product complementary to the first RNA product, where the second primer extension product has a 3'-end which is complementary to the 5'-end of the first RNA product; separating the second primer extension product from the first RNA product using an enzyme which selectively degrades the first RNA product, *e.g.*, an enzyme with an RNAse H activity; treating the second primer extension product with the promoter oligonucleotide described above to form a promoter oligonucleotide:second primer extension product hybrid; extending the 3'-end of the second primer extension product in the promoter oligonucleotide:second primer extension product hybrid to add a sequence complementary to the second region of the promoter oligonucleotide; and transcribing from the promoter oligonucleotide:second primer extension product hybrid multiple second RNA products complementary to the second primer extension product using an RNA polymerase, where the base sequences of the second RNA products are substantially identical to the base sequence of the target sequence.

Another aspect of the method of this embodiment comprises treating the target nucleic acid with a binding molecule (*e.g.*, terminating oligonucleotide or digestion oligonucleotide), where the binding molecule binds to the target nucleic acid adjacent to or near the 5'-end of the target sequence (as indicated above, the phrase "adjacent to" means that the binding molecule binds to a base sequence of the target nucleic acid next to the 5'-terminal base of the target sequence and fully 5' to the target sequence). The target nucleic acid is preferably treated with the binding molecule prior to initiating extension of the priming oligonucleotide in a primer extension reaction with a DNA polymerase. In this aspect, the first primer extension product has a 3'-end which is determined by the binding molecule, where the 3'-end of the primer extension product is complementary to the 5'-end of the target sequence. After the promoter oligonucleotides hybridizes to the first primer extension product, this aspect of the method further comprises extending the 3'-end of the first DNA primer extension product in the promoter oligonucleotide:first primer extension product hybrid to add a sequence complementary to the second region of the promoter oligonucleotide.

Yet another aspect of the method of this embodiment comprises treating the target nucleic acid with a displacer oligonucleotide, where the displacer oligonucleotide hybridizes to the target nucleic acid upstream from the priming oligonucleotide such that a primer extension reaction can be initiated therefrom (in this context, the term "upstream" means that the 3'-terminal nucleotide of the displacer oligonucleotide is 5' to the 3'-terminal nucleotide of the priming oligonucleotide), and then extending the displacer oligonucleotide in a primer extension reaction with a DNA polymerase to give a second DNA primer extension product that displaces the first DNA primer extension product from the target nucleic acid. In a preferred aspect, the activity of the DNA polymerase is substantially limited to the formation of primer extension products comprising the displacer and priming oligonucleotides.

Reagents and conditions suitable for practicing any of the embodiments described above are set forth in the Examples section.

The methods of the present invention may be used as a component of assays to detect and/or quantitate specific nucleic acid target sequences in clinical, food, water, industrial, environmental, forensic, and similar samples or to produce large numbers of copies of DNA and/or RNA of specific target sequences for a variety of uses. (As used herein, the term "copies" refers to amplification products having either the same or the opposite sense of the target sequence.) These methods may also be used to produce multiple copies of a target sequence for cloning or to generate probes or to produce RNA and DNA copies for sequencing.

The priming oligonucleotides of the embodiments described above optionally have a cap comprising a base region hybridized to a 3'-end thereof prior to treating a target nucleic acid or an RNA product with one of the priming oligonucleotides in order to prevent the initiation of DNA synthesis therefrom. (As used herein, the term "priming oligonucleotide" is inclusive of displacer oligonucleotides.) The 5'-terminal base (*i.e.*, the 5'-most base) of a cap hybridizes to the 3'-terminal base (*i.e.*, the 3'-most base) of a priming oligonucleotide. However, the caps are designed to be preferentially displaced from priming oligonucleotides by a target nucleic acid, a primer extension product, or an RNA product. A cap of the present invention may take the form of a discrete capping oligonucleotide, or may be attached to the 5'-end of a priming oligonucleotide via a linker. A preferred capping oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, comprises a blocking moiety at its 3'-terminus).

To increase the binding affinity of a priming oligonucleotide for a target nucleic acid or complement thereof, the 5'-end of a priming oligonucleotide may include one or more modifications which improve the binding properties (*e.g.*, hybridization or base stacking) of the priming oligonucleotide to the target nucleic acid or an RNA product, provided the

modifications do not prevent the priming oligonucleotide from being extended in a primer extension reaction or substantially interfere with cleavage of an RNA template to which the priming oligonucleotide is hybridized. The modifications are preferably spaced at least 15 bases from the 3'-terminus of a priming oligonucleotide, and most preferably affect a region limited to the three or four 5'-most nucleotides of the priming oligonucleotide. Preferred modifications include 2'-O-methyl ribonucleotides and "Locked Nucleic Acids" or "Locked Nucleoside Analogues" (LNAs). *See* Becker *et al.*, "Method for Amplifying Target Nucleic Acids Using Modified Primers," U.S. Patent No. 6,13,038; Imanishi *et al.*, "Bicyclonucleoside and Oligonucleotide Analogues," U.S. Patent No. 6,268,490; and Wengel *et al.*, "Oligonucleotide Analogues," U.S. Patent No. 6,670,461. The contents of each of the foregoing references are hereby incorporated by reference herein.

The promoter oligonucleotide used in the methods described above may further include an insertion sequence which is selected to enhance the rate at which RNA products are formed. The insertion sequence is preferably from 5 to 20 nucleotides in length and is positioned between or adjacent to the first and second regions of the promoter oligonucleotide. Preferred insertion sequences of the present invention include the base sequences of SEQ ID NO:1 ccacaa and SEQ ID NO:2 acgtagcatcc.

The rate of amplification may also be affected by the inclusion of an extender oligonucleotide in any of the above-described methods. An extender oligonucleotide is preferably from 10 to 50 nucleotides in length and is designed to hybridize to a DNA template so that the 5'-end of the extender oligonucleotide is adjacent to or near the 3'-end of a promoter oligonucleotide. The extender oligonucleotide is preferably modified to prevent the initiation of DNA synthesis therefrom (*e.g.*, includes a 3'-terminal blocking moiety).

In some applications of the methods described above, the binding molecule may comprise an oligonucleotide having a 5'-end which overlaps the 5'-end of the first region of the promoter oligonucleotide. To limit hybridization of the binding molecule to the promoter oligonucleotide, the 5'-end of the first region of the promoter oligonucleotide may be synthesized to include a sufficient number of mismatches with the 5'-end of the binding molecule to prevent the promoter oligonucleotide from hybridizing to the binding molecule. While a single mismatch generally should be sufficient, the number of destabilizing mismatches needed in the first region of the promoter oligonucleotide will depend upon the length and base composition of the overlapping region.

In an adaptation of the above methods, the blocking moiety may be released from the promoter oligonucleotide prior to treating the primer extension product or the first primer extension product with the promoter oligonucleotide. To facilitate release of the blocking moiety, the promoter oligonucleotide is provided to a reaction mixture pre-hybridized to an

oligodeoxynucleotide. The oligodeoxynucleotide is hybridized to a 3'-region of the first region of the promoter oligonucleotide which includes a sufficient number of contiguous ribonucleotides such that the blocking moiety is released from the promoter oligonucleotide in the presence of an enzymatic activity capable of cleaving the ribonucleotides of the 3'-region. During cleavage of the ribonucleotides, the oligodeoxynucleotide is also released from the first region of the promoter oligonucleotide, and the remaining, uncleaved portion of the first region hybridizes to the primer extension product or the first primer extension product. The 3'-section of ribonucleotides preferably includes at least 6 contiguous ribonucleotides, and the oligdeoxyonucleotide is preferably the same length as and fully complementary to the 3'-section of ribonucleotides. The oligodeoxynucleotide may be a separate molecule or it may be joined to the promoter oligonucleotide by means of a linker.

The present invention further relates to reaction mixtures useful for carrying out the methods described above. The reaction mixtures of the present invention may contain each component, or some subcombination of components, necessary for carrying out the methods described above.

The materials and/or reagents used in the methods of the present invention may be incorporated as parts of kits, *e.g.*, diagnostic kits for clinical or criminal laboratories, or nucleic amplification kits for general laboratory use. The present invention thus includes kits which include some or all of the components necessary to carry out the methods of the present invention, *e.g.*, oligonucleotides, binding molecules, stock solutions, enzymes, positive and negative control target sequences, detection reagents, containers (*e.g.*, test tubes, cuvettes, cassettes, plates, microfluidic devices and the like), and instructions provided in written or electronic form for performing the disclosed methods.

Certain embodiments of the present invention include one or more detection probes for determining the presence or amount of the RNA and/or DNA products in the amplification reaction mixture. Probes may be designed to detect RNA and/or DNA products after the amplification reaction (*i.e.*, end-point detection) or, alternatively, during the amplification reaction (*i.e.*, real-time detection involves periodically measuring the amount of signal associated with probe:amplicon complexes in the reaction mixture). Thus, the probes may be provided to the reaction mixture prior to, during or at the completion of the amplification reaction. For real-time detection of RNA products in the first two methods described above, it may be desirable to provide the probe to the reaction mixture after the first primer extension reaction has been initiated (*i.e.*, addition of amplification enzymes) since probe binding to the target sequence, rather than RNA product, may slow the rate at which an RNA-dependent DNA polymerase (*e.g.*, reverse transcriptase) can extend the priming oligonucleotide. Preferred probes have one or more associated labels to facilitate detection.

The present invention is further drawn to various oligonucleotides, including the priming oligonucleotides, promoter oligonucleotides, terminating oligonucleotides,displacer oligonucleotides, capping oligonucleotides, extender oligonucleotides and detection probes described herein. It is to be understood that oligonucleotides of the present invention may be DNA or RNA (and analogs thereof), and in either case, the present invention includes RNA equivalents of DNA oligonucleotides and DNA equivalents of RNA oligonucleotides. Except for the preferred priming oligonucleotides, displacer oligonucleotides and detection probes described below, the oligonucleotides described in the following paragraphs are preferably modified to prevent their participation in a synthesis reaction in the presence of a DNA polymerase (*e.g.*, include a blocking moiety at their 3'-termini).

For certain amplification reactions in which the target nucleic acid contains a hepatitis C virus (HCV) 5' untranslated region, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3 aatttaatacgactcac tatagggaga. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 15, 20, 25, 30 or 32 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:4 ctagccatggcgttagtatgagtgtcgtgcag or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:5 aatttaatacgactcactatagggagactagccatggcgttagtatgagtgtcgtgcag or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25, 30 or 32 contiguous bases of the base sequence of SEQ ID NO:4 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains an HCV 5' untranslated region, the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide

comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20, 25, 30 or 31 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:6 aggcattgagcgggttgatccaagaaaggac or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:6 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes under amplification conditions to the target nucleic acid. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20, 25, 30 or 31 contiguous bases of the base sequence of SEQ ID NO:6 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains an HCV 5' untranslated region, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 13 or 15 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:7 guacucaccgguucc, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not human nucleic acid) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:7, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 10, 13 or 15 contiguous bases of the base sequence of SEQ ID NO:7, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains an HCV 5' untranslated region, the present invention is further directed to a detection probe up to 40

or 50 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 18, 20 or 22 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:8 agaccacuauggcucucccggg, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not human nucleic acid) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:8, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 18, 20 or 22 contiguous bases of the base sequence SEQ ID NO:8, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains a human immunodeficiency virus (HIV) *pol* gene, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 15, 20, 25, 30 or 31contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:9 acaaatggcagtattcatccacaatttaaaa or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:10 aatttaatacgactcactatagggagacta gccatggcgttagtatgagtgtcgtgcag or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the

target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25, 30 or 31 contiguous bases of the base sequence of SEQ ID NO:9 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains an HIV *pol* gene, the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20, 25 or 27 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:11 gtttgtatgtctgttgctattatgtct or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:11 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20, 25 or 27 contiguous bases of the base sequence of SEQ ID NO:11 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains an HIV *pol* gene, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 13, 15 or 17 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:12 acuguacccccaaucc, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not human nucleic acid) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:12, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and

which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 13, 15 or 17 contiguous bases of the base sequence of SEQ ID NO:12, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes under stringent hybridization conditions to the target nucleic acid or its complement. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains a human papilloma virus (HPV) E6 and E7 gene, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 15, 20, 25 or 27 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:13 gaacagatggggcacacaattcctagt or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:14 aatttaatacgactcactatagggagagaa cagatggggcacacaattcctagt or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25 or 27 contiguous bases of the base sequence of SEQ ID NO:13 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains an HPV E6 and E7 gene, the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15 or 19 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:15 gacagctcagaggaggagg or an equivalent

sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:15 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15 or 19 contiguous bases of the base sequence of SEQ ID NO:15 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains an HPV E6 and E7 gene, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 15, 17 or 19 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:16 ggacaagcagaaccggaca or the complement thereof, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not human nucleic acid) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:16 or the complement thereof, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 15, 17 or 19 contiguous bases of the base sequence of SEQ ID NO:16 or the complement thereof, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains a West Nile Virus (WNV) nonstructural protein 5 gene, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained

within and includes at least 10, 15, 20, 25 or 27 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:17 gagtagacggtgctgcctgcgactcaa or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:18 aatttaatacgactcactcactatagggagagagtagacggtgctgcctgcgactcaa or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25 or 27 contiguous bases of the base sequence of SEQ ID NO:17 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a WNV nonstructural protein 5 gene, the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20 or 23 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:19 tccgagacggttctgagggctta or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:19 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20 or 23 contiguous bases of the base sequence of SEQ ID NO:19 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a WNV nonstructural protein 5 gene, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and

includes at least 14, 16 or 18 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:20 gaucacuucgcggcuuug, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not human nucleic acid) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:20, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 14, 16 or 18 contiguous bases of the base sequence of SEQ ID NO:20, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains a 23S rRNA sequence of *Chlamydia trachomatis*, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 15, 20, 25 or 30 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:21 cggagtaagttaagcacgcggacgattgga or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:22 aatttaatacgactcactatagggagacgg agtaagttaagcacgcggacgattgga or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence

consisting of or contained within and including at least 10, 15, 20, 25 or 30 contiguous bases of the base sequence of SEQ ID NO:21 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a 23S rRNA sequence of *Chlamydia trachomatis*, the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20, 25 or 29 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:23 cccgaagattccccttgatcgcgacctga or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:23 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20, 25 or 29 contiguous bases of the base sequence of SEQ ID NO:23 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a 23S rRNA sequence of *Chlamydia trachomatis*, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 19, 22 or 24 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:24 cguucucaucgcucuacggacucu, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not *Chlamydia psittaci* nucleic acid) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:24, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes under stringent hybridization conditions to the target nucleic acid or its complement. The base sequence of

the detection probe preferably consists of or is contained within and includes at least 19, 22 or 24 contiguous bases of the base sequence of SEQ ID NO:24, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains a 16S rRNA sequence of *Mycobacterium tuberculosis*, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 15, 20, 25, 30, 35 or 36 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:25 actgggtctaataccggataggaccacgggatgcat or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:26 aattctaatacgactcactatagggagaactgggtctaataccggataggaccacgggatgcat or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25, 30, 35 or 36 contiguous bases of the base sequence of SEQ ID NO:25 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a 16S rRNA sequence of *Mycobacterium tuberculosis*, the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained

within and includes at least 10, 15, 20, 25, 30 or 31 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:27 actgggtctaataccggataggaccacggga or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:28 aattctaatacgactcactat agggagaactgggtctaataccggataggaccacggga or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25, 30 or 31 contiguous bases of the base sequence of SEQ ID NO:27 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a 16S rRNA sequence of *Mycobacterium tuberculosis*, the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20, 25 or 27 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:29 gccgtcaccccaccaacaagctgatag or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:29 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20, 25 or 27 contiguous bases of the base sequence of SEQ ID NO:29 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains a 16S rRNA sequence of *Mycobacterium tuberculosis*, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with,

or is contained within and includes at least 18, 20 or 22 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:30 gcucaucccacaccgcuaaagc, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not nucleic acid from a *Mycobacterium avium* complex organism) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:30, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 18, 20 or 22 contiguous bases of the base sequence of SEQ ID NO:30, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains a 16S rRNA sequence of *Mycobacterium tuberculosis*, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 22, 25 or 28 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:31 ccgagaucccacaccgcuaaagccucgg, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not nucleic acid from a *Mycobacterium avium* complex organism) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:31, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 22, 25 or 28 contiguous bases of the base sequence

of SEQ ID NO:31, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, a fluorophore/quencher dye pair.

For certain amplification reactions in which the target nucleic acid contains a 16S rRNA sequence of *Mycobacterium tuberculosis*, the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 18, 20 or 22 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:32 gctcatcccacaccgctaaagc, the complement thereof, or an equivalent sequence containing uracil bases substituted for thymine bases, and which preferentially hybridizes to the target nucleic acid or its complement (*e.g.*, not nucleic acid from a *Mycobacterium avium* complex organism) under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:32, the complement thereof, or an equivalent sequence containing uracil bases substituted for thymine bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The base sequence of the detection probe preferably consists of or is contained within and includes at least 18, 20 or 22 contiguous bases of the base sequence of SEQ ID NO:32, the complement thereof, or an equivalent sequence containing uracil bases substituted for thymine bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, an AE substituent.

For certain amplification reactions in which the target nucleic acid contains the orfX gene of a *Staphylococcus aureus* (*e.g.* a methicillin-resistant strain), the present invention includes a promoter oligonucleotide comprising a promoter sequence and a hybridizing sequence up to 40 or 50 bases in length. The promoter sequence is recognized by an RNA polymerase, such as a T7, T3 or SP6 RNA polymerase, and preferably includes the T7 RNA polymerase promoter sequence of SEQ ID NO:3 aatttaatacgactcactatagggaga. The hybridizing sequence of the preferred promoter oligonucleotide comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 15, 20, 25 or 29 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:33 tgacccaagggcaaagcgactttg or an equivalent sequence containing uracil bases

substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The promoter oligonucleotide preferably does not include a region in addition to the hybridizing sequence that hybridizes to the target nucleic acid under amplification conditions. More preferably, the promoter oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:34 aatttaatacgactcactatagggagatgacccaagggcaaagcgactttg or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. The base sequence of the promoter oligonucleotide preferably consists of a promoter sequence, such as SEQ ID NO:3, and a hybridizing sequence consisting of or contained within and including at least 10, 15, 20, 25 or 29 contiguous bases of the base sequence of SEQ ID NO:33 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains the orfX gene of a *Staphylococcus aureus* (*e.g.* a methicillin-resistant strain), the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20, 25 or 26 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:35 gtgcgtagttactgcgttgtaagacgtc or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:35 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes under amplification conditions to the target nucleic acid. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20, 25 or 26 contiguous bases of the base sequence of SEQ ID NO:35 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains the orfX gene of a *Staphylococcus aureus* (*e.g.* a methicillin-resistant strain), the present invention includes a priming oligonucleotide up to 40 or 50 bases in length. A preferred priming oligonucleotide includes an oligonucleotide comprising, consisting of, consisting essentially of, overlapping with, or contained within and including at least 10, 15, 20, 25 or 26 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:36 ctgaatgatagtgcgtagttactgcg or an equivalent sequence containing

uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions. More preferably, the priming oligonucleotide comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:36 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes under amplification conditions to the target nucleic acid. The base sequence of the priming oligonucleotide preferably consists of or is contained within and includes at least 10, 15, 20, 25 or 26 contiguous bases of the base sequence of SEQ ID NO:36 or an equivalent sequence containing uracil bases substituted for thymine bases, and which hybridizes to the target nucleic acid under amplification conditions.

For certain amplification reactions in which the target nucleic acid contains the orfX gene of a *Staphylococcus aureus* (*e.g.* a methicillin-resistant strain), the present invention is further directed to a detection probe up to 35, 50 or 100 bases in length. A preferred detection probe includes a target binding region which comprises, consists of, consists essentially of, overlaps with, or is contained within and includes at least 10, 12, 15 or 17 contiguous bases of a base sequence that is at least 80%, 90% or 100% identical to the base sequence of SEQ ID NO:37 ccgucauuggcggauca, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The detection probe preferably does not include a region in addition to the target binding region that hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. More preferably, the base sequence of the target binding region of the detection probe comprises, consists of, or consists essentially of a base sequence substantially corresponding to the base sequence of SEQ ID NO:38, the complement thereof, or an equivalent sequence containing thymine bases substituted for uracil bases, and which preferentially hybridizes to the target nucleic acid or its complement under stringent hybridization conditions. The detection probe may be capable of forming a hairpin molecule through self-hybridization at its end portions, such as a "molecular beacon" or "molecular torch," as described *infra*, under the stringent hybridization conditions. In certain embodiments the probe optionally includes one or more detectable labels, *e.g.*, a pair of interacting fluorescent labels.

For amplification reactions which do not form part of the present invention, the above-described promoter oligonucleotides may be modified to exclude the promoter sequence and/or the priming oligonucleotides may be modified to include a promoter sequence. Additionally, where the desired specificity for a target sequence can be achieved, the promoter oligonucleotides and/or the priming oligonucleotides described above may be modified and used as detection probes. Also, the above-described detection probes may be adapted for use as amplification oligonucleotides.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1D depict four general methods of the present invention.

Figures 2A-2D depict the general methods of Figures 1A-1D with the further inclusion of an extender oligonucleotide hybridized to an extension product or target sequence 3' to the blocked promoter oligonucleotide.

FIG. 3 depicts a denaturing agarose gel showing the effect of using a promoter oligonucleotide with a 3'-blocking moiety.

FIG. 4 shows the real-time accumulation of amplification products in a *Mycobacterium tuberculosis* system, both in the presence (Figures 4A, 4C and 4E) and in the absence (Figures 4B, 4D and 4F) of a terminating oligonucleotide modified to fully contain 2'-O-methyl ribonucleotides. The input target nucleic acid for these reactions was 0 copies (Figures 4A and 4B), 100 copies (Figures 4C and 4D) and 1000 copies (Figures 4E and 4F).

FIG. 5 illustrates the formation of primer-dependent side-products.

Figures 6A and 6B illustrate the use of caps to limit side-product formation. The cap and priming oligonucleotide are separate molecules in FIG. 6A, and in FIG. 6B they are linked to each other.

Figures 7A and 7B depict non-denaturing agarose gels showing the effect of a capping oligonucleotide on side-product formation. FIG. 7A depicts reactions without added template, and FIG. 7B depicts reactions with added template.

FIG. 8 shows the real-time accumulation of amplification products in a *Staphylococcus aureus* system according to a method of the present invention. The target nucleic acid was double-stranded DNA, which was tested at four copy levels against a negative control.

FIG. 9 shows the real-time accumulation of amplification products in five different *Staphylococcus aureus* systems according to methods of the present invention. The target nucleic acid for each system was double-stranded DNA, and each was system was tested at 10,000 copies of target against a negative control. The first system (FIG. 9A) employed the method of FIG. 1D, and each of the remaining systems excluded a component of this method. The method of FIG. 9B excluded the terminating oligonucleotide; the method of FIG. 9C excluded the displacer oligonucleotide; the method of FIG. 9D excluded the priming oligonucleotide; and the method of FIG. 9E excluded a denaturation step.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, novel methods, reaction mixtures, compositions and kits are provided for the amplification of specific nucleic acid target sequences for use in assays for the detection and/or quantitation of such nucleic acid target sequences or for the production of large numbers of copies of DNA and/or RNA of specific target sequences for a variety of uses. In particular, the embodiments of the present invention provide for amplification of nucleic acid target sequences with enhanced specificity and sensitivity. Amplification methods of the present invention are preferably carried out using priming oligonucleotides that target only one sense of a target nucleic acid, with all other oligonucleotides used in the amplification methods preferably comprising a blocking moiety at their 3'-termini so that they cannot be extended by a nucleic acid polymerase.

Definitions

The following terms have the following meanings unless expressly stated to the contrary. It is to be noted that the term "a" or "an" entity refers to one or more of that entity; for example, "a nucleic acid," is understood to represent one or more nucleic acids. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

1. Nucleic acid

The term "nucleic acid" is intended to encompass a singular "nucleic acid" as well as plural "nucleic acids," and refers to any chain of two or more nucleotides, nucleosides, or nucleobases (*e.g.*, deoxyribonucleotides or ribonucleotides) covalently bonded together. Nucleic acids include, but are not limited to, virus genomes, or portions thereof, either DNA or RNA, bacterial genomes, or portions thereof, fungal, plant or animal genomes, or portions thereof, messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), plasmid DNA, mitochondrial DNA, or synthetic DNA or RNA. A nucleic acid may be provided in a linear (*e.g.*, mRNA), circular (*e.g.*, plasmid), or branched form, as well as a double-stranded or single-stranded form. Nucleic acids may include modified bases to alter the function or behavior of the nucleic acid, *e.g.*, addition of a 3'-terminal dideoxynucleotide to block additional nucleotides from being added to the nucleic acid. As used herein, a "sequence" of a nucleic acid refers to the sequence of bases which make up a nucleic acid. The term "polynucleotide" may be used herein to denote a nucleic acid chain. Throughout this application, nucleic acids are designated as having a 5'-terminus and a 3'-terminus. Standard nucleic acids, *e.g.*, DNA and RNA, are typically synthesized "5'-to-3'," *i.e.*, by the addition of nucleotides to the 3'-terminus of a growing nucleic acid.

A "nucleotide" is a subunit of a nucleic acid consisting of a phosphate group, a 5-carbon sugar and a nitrogenous base. The 5-carbon sugar found in RNA is ribose. In DNA, the 5-carbon sugar is 2'-deoxyribose. The term also includes analogs of such subunits, such as a methoxy group at the 2' position of the ribose (2'-O-Me). As used herein, methoxy oligonucleotides containing "T" residues have a methoxy group at the 2' position of the ribose moiety, and a uracil at the base position of the nucleotide.

A "non-nucleotide unit" is a unit which does not significantly participate in hybridization of a polymer. Such units must not, for example, participate in any significant hydrogen bonding with a nucleotide, and would exclude units having as a component one of the five nucleotide bases or analogs thereof.

2. Oligonucleotide

As used herein, the term "oligonucleotide" or "oligomer" is intended to encompass a singular "oligonucleotide" as well as plural "oligonucleotides," and refers to any polymer of two or more of nucleotides, nucleosides, nucleobases or related compounds used as a reagent in the amplification methods of the present invention, as well as subsequent detection methods. The oligonucleotide may be DNA and/or RNA and/or analogs thereof. The term oligonucleotide does not denote any particular function to the reagent, rather, it is used generically to cover all such reagents described herein. An oligonucleotide may serve various different functions, *e.g.*, it may function as a primer if it is capable of hybridizing to a complementary strand and can further be extended in the presence of a nucleic acid polymerase, it may provide a promoter if it contains a sequence recognized by an RNA polymerase and allows for transcription, and it may function to prevent hybridization or impede primer extension if appropriately situated and/or modified. Specific oligonucleotides of the present invention are described in more detail below. As used herein, an oligonucleotide can be virtually any length, limited only by its specific function in the amplification reaction or in detecting an amplification product of the amplification reaction. Oligonucleotides of a defined sequence and chemical structure may be produced by techniques known to those of ordinary skill in the art, such as by chemical or biochemical synthesis, and by *in vitro* or *in vivo* expression from recombinant nucleic acid molecules, *e.g.*, bacterial or viral vectors. As intended by this disclosure, an oligonucleotide does not consist solely of wild-type chromosomal DNA or the *in vivo* transcription products thereof.

Oligonucleotides may be modified in any way, as long as a given modification is compatible with the desired function of a given oligonucleotide. One of ordinary skill in the art can easily determine whether a given modification is suitable or desired for any given oligonucleotide of the present invention. Modifications include base modifications, sugar

modifications or backbone modifications. Base modifications include, but are not limited to the use of the following bases in addition to adenine, cytidine, guanosine, thymine and uracil: C-5 propyne, 2-amino adenine, 5-methyl cytidine, inosine, and dP and dK bases. The sugar groups of the nucleoside subunits may be ribose, deoxyribose and analogs thereof, including, for example, ribonucleosides having a 2'-O-methyl substitution to the ribofuranosyl moiety. *See* Becker *et al.*, U.S. Patent No. 6,130,038. Other sugar modifications include, but are not limited to 2'-amino, 2'-fluoro, (L)-alpha-threofuranosyl, and pentopuranosyl modifications. The nucleoside subunits may by joined by linkages such as phosphodiester linkages, modified linkages or by non-nucleotide moieties which do not prevent hybridization of the oligonucleotide to its complementary target nucleic acid sequence. Modified linkages include those linkages in which a standard phosphodiester linkage is replaced with a different linkage, such as a phosphorothioate linkage or a methylphosphonate linkage. The nucleobase subunits may be joined, for example, by replacing the natural deoxyribose phosphate backbone of DNA with a pseudo peptide backbone, such as a 2-aminoethylglycine backbone which couples the nucleobase subunits by means of a carboxymethyl linker to the central secondary amine. (DNA analogs having a pseudo peptide backbone are referred to as "peptide nucleic acids" or "PNA" and disclosed by Nielsen *et al.*, "Peptide Nucleic Acids," U.S. Patent No. 5,539,082.) Other linkage modifications include, but are not limited to, morpholino bonds.

Non-limiting examples of oligonucleotides or oligomers contemplated by the present invention include nucleic acid analogs containing bicyclic and tricyclic nucleoside and nucleotide analogs (LNAs). *See* Imanishi *et al.*, U.S. Patent No. 6,268,490; and Wengel *et al.*, U.S. Patent No. 6,670,461.) Any nucleic acid analog is contemplated by the present invention provided the modified oligonucleotide can perform its intended function, *e.g.*, hybridize to a target nucleic acid under stringent hybridization conditions or amplification conditions, or interact with a DNA or RNA polymerase, thereby initiating extension or transcription. In the case of detection probes, the modified oligonucleotides must also be capable of preferentially hybridizing to the target nucleic acid under stringent hybridization conditions.

While design and sequence of oligonucleotides for the present invention depend on their function as described below, several variables must generally be taken into account. Among the most critical are: length, melting temperature (Tm), specificity, complementarity with other oligonucleotides in the system, G/C content, polypyrimidine (T, C) or polypurine (A, G) stretches, and the 3'-end sequence. Controlling for these and other variables is a standard and well known aspect of oligonucleotide design, and various computer programs are readily available to screen large numbers of potential oligonucleotides for optimal ones.

The 3'-terminus of an oligonucleotide (or other nucleic acid) can be blocked in a variety of ways using a blocking moiety, as described below. A "blocked" oligonucleotide

cannot be extended by the addition of nucleotides to its 3'-terminus, by a DNA- or RNA-dependent DNA polymerase, to produce a complementary strand of DNA. As such, a "blocked" oligonucleotide cannot be a "priming oligonucleotide."

As used in this disclosure, the phrase "an oligonucleotide having a nucleic acid sequence 'comprising,' 'consisting of,' or 'consisting essentially of' a sequence selected from" a group of specific sequences means that the oligonucléotide, as a basic and novel characteristic, is capable of stably hybridizing to a nucleic acid having the exact complement of one of the listed nucleic acid sequences of the group under stringent hybridization conditions. An exact complement includes the corresponding DNA or RNA sequence.

The phrase "an oligonucleotide substantially corresponding to a nucleic acid sequence" means that the referred to oligonucleotide is sufficiently similar to the reference nucleic acid sequence such that the oligonucleotide has similar hybridization properties to the reference nucleic acid sequence in that it would hybridize with the same target nucleic acid sequence under stringent hybridization conditions.

One skilled in the art will understand that "substantially corresponding" oligonucleotides of the invention can vary from the referred to sequence and still hybridize to the same target nucleic acid sequence. This variation from the nucleic acid may be stated in terms of a percentage of identical bases within the sequence or the percentage of perfectly complementary bases between the probe or primer and its target sequence. Thus, an oligonucleotide of the present invention substantially corresponds to a reference nucleic acid sequence if these percentages of base identity or complementarity are from 100% to about 80%. In preferred embodiments, the percentage is from 100% to about 85%. In more preferred embodiments, this percentage can be from 100% to about 90%; in other preferred embodiments, this percentage is from 100% to about 95%. One skilled in the art will understand the various modifications to the hybridization conditions that might be required at various percentages of complementarity to allow hybridization to a specific target sequence without causing an unacceptable level of non-specific hybridization.

3. Blocking Moiety

As used herein, a "blocking moiety" is a substance used to "block" the 3'-terminus of an oligonucleotide or other nucleic acid so that it cannot be extended by a nucleic acid polymerase. A blocking moiety may be a small molecule, *e.g.*, a phosphate or ammonium group, or it may be a modified nucleotide, *e.g.*, a 3'2' dideoxynucleotide or 3' deoxyadenosine 5'-triphosphate (cordycepin), or other modified nucleotide. Additional blocking moieties include, for example, the use of a nucleotide or a short nucleotide sequence having a 3'-to-5' orientation, so that there is no free hydroxyl group at the 3'-terminus, the use of a 3' alkyl

group, a 3' non-nucleotide moiety (*see, e.g.*, Arnold *et al.*, "Non-Nucleotide Linking Reagents for Nucleotide Probes," U.S. Patent No. 6,031,091, the contents of which are hereby incorporated by reference herein), phosphorothioate, alkane-diol residues, peptide nucleic acid (PNA), nucleotide residues lacking a 3' hydroxyl group at the 3'-terminus, or a nucleic acid binding protein. Preferably, the 3'-blocking moiety comprises a nucleotide or a nucleotide sequence having a 3'-to-5' orientation or a 3' non-nucleotide moiety, and not a 3'2'-dideoxynucleotide or a 3' terminus having a free hydroxyl group. Additional methods to prepare 3'-blocking oligonucleotides are well known to those of ordinary skill in the art.

4. Binding molecule

As used herein, a "binding molecule" is a substance which hybridizes to or otherwise binds to a target nucleic acid adjacent to or near the 5'-end of the desired target sequence, so as to limit a DNA primer extension product to a desired length, *i.e.*, a primer extension product having a generally defined 3'-end. As used herein, the phrase "defined 3'-end" means that the 3'-end of a primer extension product is not wholly indeterminate, as would be the case in a primer extension reaction which occurs in the absence of a binding molecule, but rather that the 3'-end of the primer extension product is generally known to within a small range of bases. In certain embodiments, a binding molecule comprises a base region. The base region may be DNA, RNA, a DNA:RNA chimeric molecule, or an analog thereof. Binding molecules comprising a base region may be modified in one or more ways, as described herein. Exemplary base regions include terminating and digestion oligonucleotides, as described below. In other embodiments, a binding molecule may comprise, for example, a protein or drug capable of binding RNA with sufficient affinity and specificity to limit a DNA primer extension product to a pre-determined length.

5. Terminating Oligonucleotide

In the present invention, a "terminating oligonucleotide" is an oligonucleotide comprising a base sequence that is complementary to a region of the target nucleic acid in the vicinity of the 5'-end of the target sequence, so as to "terminate" primer extension of a nascent nucleic acid that includes a priming oligonucleotide, thereby providing a defined 3'-end for the nascent nucleic acid strand. A terminating oligonucleotide is designed to hybridize to the target nucleic acid at a position sufficient to achieve the desired 3'-end for the nascent nucleic acid strand. The positioning of the terminating oligonucleotide is flexible depending upon its design. A terminating oligonucleotide may be modified or unmodified. In certain embodiments, terminating oligonucleotides are synthesized with at least one or more 2'-O-methyl ribonucleotides. These modified nucleotides have demonstrated higher thermal

stability of complementary duplexes. The 2'-O-methyl ribonucleotides also function to increase the resistance of oligonucleotides to exonucleases, thereby increasing the half-life of the modified oligonucleotides. *See, e.g.*, Majlessi *et al.* (1988) *Nucleic Acids Res.* 26, 2224-9, the contents of which are hereby incorporated by reference herein. Other modifications as described elsewhere herein may be utilized in addition to or in place of 2'-O-methyl ribonucleotides. For example, a terminating oligonucleotide may comprise PNA or an LNA. *See, e.g.*, Petersen *et al.* (2000) *J. Mol. Recognit.* 13, 44-53, the contents of which are hereby incorporated by reference herein. Although not required, a terminating oligonucleotide of the present invention preferably includes a blocking moiety at its 3'-terminus to prevent extension. A terminating oligonucleotide may also comprise a protein or peptide joined to the oligonucleotide so as to terminate further extension of a nascent nucleic acid chain by a polymerase. A terminating oligonucleotide of the present invention is typically at least 10 bases in length, and may extend up to 15, 20, 25, 30, 35, 40, 50 or more nucleotides in length. Suitable and preferred terminating oligonucleotides are described herein. It should be noted that while a terminating oligonucleotide typically or necessarily includes a 3'-blocking moiety, "3'-blocked" oligonucleotides are not necessarily terminating oligonucleotides. Other oligonucleotides of the present invention, *e.g.*, promoter oligonucleotides and capping oligonucleotides are typically or necessarily 3'-blocked as well.

6. Modifying Oligonucleotide/Digestion Oligonucleotide

A modifying oligonucleotide provides a mechanism by which the 3'-terminus of the primer extension product is determined. A modifying oligonucleotide typically comprises a motif which hybridizes to one or more bases in the vicinity of the 5'-end of a target sequence, and which facilitates termination of primer extension by means of a modifying enzyme, *e.g.*, a nuclease. Alternatively, a modifying oligonucleotide might comprise a base region which hybridizes in the vicinity of the 3'-end of a target sequence, and is tethered to a specific modifying enzyme or to a chemical which can then terminate primer extension.

One specific modifying oligonucleotide is a digestion oligonucleotide. A digestion oligonucleotide is comprised of DNA, preferably a stretch of at least about 6 deoxyribonucleotides. The digestion oligonucleotide hybridizes to the RNA template, and the RNA of a RNA:DNA hybrid is digested by a selective RNAse as described herein, *e.g.*, by an enzyme having an RNAse H activity.

7. Promoter Oligonucleotide/Promoter Sequence

As is well known in the art, a "promoter" is a specific nucleic acid sequence that is recognized by a DNA-dependent RNA polymerase ("transcriptase") as a signal to bind to the

nucleic acid and begin the transcription of RNA at a specific site. For binding, it was generally thought that such transcriptases required DNA which had been rendered double-stranded in the region comprising the promoter sequence via an extension reaction, however, the present inventors have determined that efficient transcription of RNA can take place even under conditions where a double-stranded promoter is not formed through an extension reaction with the template nucleic acid. The template nucleic acid (the sequence to be transcribed) need not be double-stranded. Individual DNA-dependent RNA polymerases recognize a variety of different promoter sequences which can vary markedly in their efficiency in promoting transcription. When an RNA polymerase binds to a promoter sequence to initiate transcription, that promoter sequence is not part of the sequence transcribed. Thus, the RNA transcripts produced thereby will not include that sequence.

According to the present invention, a "promoter oligonucleotide" refers to an oligonucleotide comprising first and second regions, and which is modified to prevent the initiation of DNA synthesis from its 3'-terminus. The "first region" of a promoter oligonucleotide of the present invention comprises a base sequence which hybridizes to a DNA template, where the hybridizing sequence is situated 3', but not necessarily adjacent to, a promoter region. The hybridizing portion of a promoter oligonucleotide of the present invention is typically at least 10 nucleotides in length, and may extend up to 15, 20, 25, 30, 35, 40, 50 or more nucleotides in length. The "second region" comprises a promoter for an RNA polymerase. A promoter oligonucleotide of the present invention is engineered so that it is incapable of being extended by an RNA- or DNA-dependent DNA polymerase, *e.g.*, reverse transcriptase, preferably comprising a blocking moiety at its 3'-terminus as described above. Suitable and preferred promoter oligonucleotides are described herein.

Promoter oligonucleotides of the present invention may be provided to a reaction mixture with an oligodeoxynucleotide bound to a ribonucleotide-containing section of the first region. The ribonucleotide-containing section preferably comprises at least 6 contiguous ribonucleotides positioned at or near the 3'-end of the first region, and the oligodeoxynucleotide is preferably the same length as and fully complementary to the ribonucleotide-containing section of the first region. Upon exposure to an enzyme capable of cleaving the RNA of an RNA:DNA duplex (*e.g.*, an RNAse H activity), a blocking moiety at the 3'-end of the promoter oligonucleotide is released and the remainder of the first region is in a single-stranded form which is available for hybridization to a DNA template. The remaining, uncleaved portion of the first region is preferably 10 to 50 nucleotides in length, as described above.

Promoter oligonucleotides of the present invention may be used in methods disclosed herein to distinguish between regions of variability in the target sequence or between the

target sequence and non-target nucleic acid which may be present in a reaction mixture (*e.g.*, single nucleotide polymorphisms ("SNPs") or mismatches between closely related organisms or strains). When a primer extension product or nucleic acid sharing a high degree of sequence identity with a target nucleic acid contains one or more base mismatches with the target binding portion of a promoter oligonucleotide, the complements of these mismatched bases are integrated into the transcription products and the mismatched bases are then reproduced in initial and/or subsequent rounds of primer extension, thereby affecting the efficiency and sensitivity of the amplification reaction. Additionally, mismatches between the target binding portion of the promoter oligonucleotide and the primer extension product may interfere with the formation of a double-stranded promoter sequence in the presence of a reverse transcriptase, thereby further affecting the efficiency of the amplification reaction. These deleterious effects on an amplification reaction can be exploited to distinguish between nucleic acids exhibiting sequence variability in the region targeted by the promoter oligonucleotide.

8. Insertion Sequence

As used herein, an "insertion sequence" is a sequence positioned between the first region (*i.e.*, template binding portion) and the second region of a promoter oligonucleotide. Insertion sequences are preferably 5 to 20 nucleotides in length, more preferably 6 to 18 nucleotides in length, and most preferably 6 to 12 nucleotides in length. The inclusion of insertion sequences in promoter oligonucleotides increases the rate at which RNA amplification products are formed. Exemplary insertion sequences are described herein.

9. Extender Oligonucleotide

An extender oligonucleotide is an oligonucleotide which hybridizes to a DNA template adjacent to or near the 3'-end of the first region of a promoter oligonucleotide. An extender oligonucleotide preferably hybridizes to a DNA template such that the 5'-terminal base of the extender oligonucleotide is within 3, 2 or 1 bases of the 3'-terminal base of a promoter oligonucleotide. Most preferably, the 5'-terminal base of an extender oligonucleotide is adjacent to the 3'-terminal base of a promoter oligonucleotide when the extender oligonucleotide and the promoter oligonucleotide are hybridized to a DNA template. To prevent extension of an extender oligonucleotide, a 3'-terminal blocking moiety is typically included. An extender oligonucleotide is preferably 10 to 50 nucleotides in length, more preferably 20 to 40 nucleotides in length, and most preferably 30 to 35 nucleotides in length.

10. Priming Oligonucleotide

A priming oligonucleotide is an oligonucleotide, at least the 3'-end of which is complementary to a nucleic acid template, and which hybridizes to the template to give a priming oligonucleotide:template hybrid suitable for initiation of synthesis by an RNA- or DNA-dependent DNA polymerase. A priming oligonucleotide is extended by the addition of covalently bonded nucleotides to its 3'-terminus, which nucleotides are complementary to the template. The result is a primer extension product. A priming oligonucleotide of the present invention is typically at least 10 nucleotides in length, and may extend up to 15, 20, 25, 30, 35, 40, 50 or more nucleotides in length. Suitable and preferred priming oligonucleotides are described herein. Virtually all DNA polymerases (including reverse transcriptases) that are known require hybridization of an oligonucleotide to a single-stranded template ("priming") to initiate DNA synthesis, whereas RNA replication and transcription (copying of RNA from DNA) generally do not require a primer. Because of its function, a priming oligonucleotide cannot comprise a 3'-blocking moiety that prevents extension in the presence of a DNA polymerase. Priming oligonucleotides are preferably designed to preferentially hybridize to a target nucleic acid, and so that it cannot be cleaved by a ribonuclease when hybridized to the target nucleic acid.

11. Displacer Oligonucleotide

A "displacer oligonucleotide" is a priming oligonucleotide which hybridizes to a template nucleic acid upstream from a neighboring priming oligonucleotide hybridized to the 3'-end of a target sequence (referred to herein as the "forward priming oligonucleotide"). By "upstream" is meant that a 3'-end of the displacer oligonucleotide complexes with the template nucleic acid 5' to a 3'-end of the forward priming oligonucleotide. When hybridized to the template nucleic acid, the 3'-terminal base of the displacer oligonucleotide is preferably adjacent to or spaced from the 5-terminal base of the forward priming oligonucleotide. More preferably, the 3'-terminal base of the displacer oligonucleotide is spaced from 5 to 35 bases from the 5'-terminal base of the forward priming oligonucleotide. The displacer oligonucleotide may be provided to a reaction mixture contemporaneously with the forward priming oligonucleotide or after the forward priming oligonucleotide has had sufficient time to hybridize to the template nucleic acid. Extension of the forward priming oligonucleotide can be initiated prior to or after the displacer oligonucleotide is provided to a reaction mixture. Under amplification conditions, the displacer oligonucleotide is extended in a template-dependent manner, thereby displacing a primer extension product comprising the forward priming oligonucleotide which is complexed with the template nucleic acid. Once displaced from the template nucleic acid, the primer extension product comprising the

forward priming oligonucleotide is available for complexing with a promoter oligonucleotide. The forward priming oligonucleotide and the displacer oligonucleotide both preferentially hybridize to the target nucleic acid.

12. Cap or Capping Oligonucleotide

As used herein, a "cap" comprises an oligonucleotide complementary to the 3'-end of a priming oligonucleotide, where the 5'-terminal base of the cap hybridizes to the 3'-terminal base of the priming oligonucleotide. A cap according to present invention is designed to preferentially hybridize to the 3'-end of the priming oligonucleotide, *e.g.*, not with a promoter oligonucleotide, but such that the cap will be displaced by hybridization of the priming oligonucleotide to the target nucleic acid. A cap may take the form of a discrete capping oligonucleotide or it may be joined to the 5'-end of the priming oligonucleotide via a linker region, thereby forming a stem-loop structure with the priming oligonucleotide under amplification conditions. Such a linker region can comprise conventional nucleotides, abasic nucleotides or otherwise modified nucleotides, or a non-nucleotide region. As described in more detail herein, a suitable cap is at least three bases in length, and is no longer than about 14 bases in length. Typical caps are about 5 to 7 bases in length.

13. Probe

By "probe" or "detection probe" is meant a molecule comprising an oligonucleotide having a base sequence partly or completely complementary to a region of an amplification product containing either sense of a target sequence sought to be detected, so as to hybridize thereto under stringent hybridization conditions. As would be understood by someone having ordinary skill in the art, a probe comprises an isolated nucleic acid molecule, or an analog thereof, in a form not found in nature without human intervention (*e.g.*, recombined with foreign nucleic acid, isolated, or purified to some extent).

The probes of this invention may have additional nucleotides outside of the targeted region so long as such nucleotides do not substantially affect hybridization under stringent hybridization conditions and, in the case of detection probes, do not prevent preferential hybridization to the target sequence. A non-complementary sequence may also be included, such as a target capture sequence (generally a homopolymer tract, such as a poly-A, poly-T

or poly-U tail), promoter sequence, a binding site for RNA transcription, a restriction endonuclease recognition site, or may contain sequences which will confer a desired secondary or tertiary structure, such as a catalytic active site or a hairpin structure on the probe, on the target nucleic acid, or both.

The probes preferably include at least one detectable label. The label may be any suitable labeling substance, including but not limited to a radioisotope, an enzyme, an enzyme cofactor, an enzyme substrate, a dye, a hapten, a chemiluminescent molecule, a fluorescent molecule, a phosphorescent molecule, an electrochemiluminescent molecule, a chromophore, a base sequence region that is unable to stably hybridize to the target nucleic acid under the stated conditions, and mixtures of these. In one particularly preferred embodiment, the label is an acridinium ester. Probes may also include interacting labels which emit different signals, depending on whether the probes have hybridized to target sequences. Examples of interacting labels include enzyme/substrates, enzyme/cofactor, luminescent/quencher, luminescent/adduct, dye dimers, and Förrester energy transfer pairs. Certain probes of the present invention do not include a label. For example, non-labeled "capture" probes may be used to enrich for target sequences or replicates thereof, which may then be detected by a second "detection" probe. *See, e.g.*, Weisburg *et al.*, "Two-Step Hybridization and Capture of a Polynucleotide," U.S. Patent No. 6,534,273, the contents of which are hereby incorporated by reference herein. While detection probes are typically labeled, certain detection technologies do not require that the probe be labeled for detection of the amplification product. *See, e.g.*, Nygren *et al.*, "Devices and Methods for Optical Detection of Nucleic Acid Hybridization, U.S. Patent No. 6,060,237.

By "stable" or "stable for detection" is meant that the temperature of a reaction mixture is at least 2°C below the melting temperature of a nucleic acid duplex. The temperature of the reaction mixture is more preferably at least 5°C below the melting temperature of the nucleic acid duplex, and even more preferably at least 10°C below the melting temperature of the nucleic acid duplex.

By "preferentially hybridize" is meant that under stringent hybridization conditions, probes of the present invention hybridize to an amplification product containing either sense of the target sequence to form stable probe:target hybrids, while at the same time formation of stable probe:non-target hybrids is minimized. Thus, a probe hybridizes to a target sequence or replicate thereof to a sufficiently greater extent than to a non-target sequence, to enable one having ordinary skill in the art to accurately quantitate the RNA replicates or complementary DNA (cDNA) of the target sequence formed during the amplification.

Probes of a defined sequence may be produced by techniques known to those of ordinary skill in the art, such as by chemical synthesis, and by *in vitro* or *in vivo* expression

from recombinant nucleic acid molecules. Preferably probes are 10 to 100 nucleotides in length, more preferably 12 to 50 bases in length, and even more preferably 18 to 35 bases in length.

14. Hybridize/Hybridization

Nucleic acid hybridization is the process by which two nucleic acid strands having completely or partially complementary nucleotide sequences come together under predetermined reaction conditions to form a stable, double-stranded hybrid. Either nucleic acid strand may be a deoxyribonucleic acid (DNA) or a ribonucleic acid (RNA) or analogs thereof. Thus, hybridization can involve RNA:RNA hybrids, DNA:DNA hybrids, RNA:DNA hybrids, or analogs thereof. The two constituent strands of this double-stranded structure, sometimes called a hybrid, are held together by hydrogen bonds. Although these hydrogen bonds most commonly form between nucleotides containing the bases adenine and thymine or uracil (A and T or U) or cytosine and guanine (C and G) on single nucleic acid strands, base pairing can also form between bases which are not members of these "canonical" pairs. Non-canonical base pairing is well-known in the art. (*See, e.g.*, ROGER L.P. ADAMS ET AL., THE BIOCHEMISTRY OF THE NUCLEIC ACIDS (11th ed. 1992).)

"Stringent hybridization conditions" or "stringent conditions" refer to conditions wherein a specific detection probe is able to hybridize to an amplification product containing either sense of a target sequence over other nucleic acids present in a reaction mixture. It will be appreciated that these conditions may vary depending upon factors including the GC content and length of the probe, the hybridization temperature, the composition of the hybridization reagent or solution, and the degree of hybridization specificity sought. Specific stringent hybridization conditions are provided in the disclosure below.

By "nucleic acid hybrid" or "hybrid" or "duplex" is meant a nucleic acid structure containing a double-stranded, hydrogen-bonded region wherein each strand is complementary to the other, and wherein the region is sufficiently stable under stringent hybridization conditions to be detected by means including, but not limited to, chemiluminescent or fluorescent light detection, autoradiography, or gel electrophoresis. Such hybrids may comprise RNA:RNA, RNA:DNA, or DNA:DNA duplex molecules.

By "complementary" is meant that the nucleotide sequences of similar regions of two single-stranded nucleic acids, or to different regions of the same single-stranded nucleic acid have a nucleotide base composition that allow the single-stranded regions to hybridize together in a stable, double-stranded hydrogen-bonded region under stringent hybridization or amplification conditions. When a contiguous sequence of nucleotides of one single-stranded region is able to form a series of "canonical" hydrogen-bonded base pairs with an

analogous sequence of nucleotides of the other single-stranded region, such that A is paired with U or T and C is paired with G, the nucleotides sequences are "perfectly" complementary.

By "preferentially hybridize" is meant that under stringent hybridization conditions, certain complementary nucleotide sequences hybridize to form a stable hybrid preferentially over other, less stable duplexes.

15. Nucleic Acid "Identity"

In certain embodiments, a nucleic acid of the present invention comprises a contiguous base region that is at least 80%, 90%, or 100% identical to a contiguous base region of a reference nucleic acid. For short nucleic acids, *e.g.*, certain oligonucleotides of the present invention, the degree of identity between a base region of a "query" nucleic acid and a base region of a reference nucleic acid can be determined by manual alignment. "Identity" is determined by comparing just the sequence of nitrogenous bases, irrespective of the sugar and backbone regions of the nucleic acids being compared. Thus, the query:reference base sequence alignment may be DNA:DNA, RNA:RNA, DNA:RNA, RNA:DNA, or any combinations or analogs thereof. Equivalent RNA and DNA base sequences can be compared by replacing uridine residues (in RNA) with thymidine residues (in DNA).

16. Target Nucleic Acid/Target Sequence

A "target nucleic acid" is a nucleic acid comprising a "target sequence" to be amplified. Target nucleic acids may be DNA or RNA as described herein, and may be either single-stranded or double-stranded. The target nucleic acid may include other sequences besides the target sequence which may not be amplified. Typical target nucleic acids include virus genomes, bacterial genomes, fungal genomes, plant genomes, animal genomes, rRNA, tRNA, or mRNA from viruses, bacteria or eukaryotic cells, mitochondrial DNA, or chromosomal DNA.

Target nucleic acids may be isolated from any number of sources based on the purpose of the amplification assay being carried out. Sources of target nucleic acids include, but are not limited to, clinical specimens, *e.g.*, blood, urine, saliva, feces, semen, or spinal fluid, from criminal evidence, from environmental samples, *e.g.*, water or soil samples, from food, from industrial samples, from cDNA libraries, or from total cellular RNA. By "isolated" it is meant that a sample containing a target nucleic acid is taken from its natural milieu, but the term does not connote any degree of purification. If necessary, target nucleic acids of the present invention are made available for interaction with the various oligonucleotides of the present invention. This may include, for example, cell lysis or cell permeabilization to release

the target nucleic acid from cells, which then may be followed by one or more purification steps, such as a series of isolation and wash steps. *See, e.g.*, Clark *et al.*, "Method for Extracting Nucleic Acids from a Wide Range of Organisms," U.S. Patent No. 5,786,208, the contents of which are hereby incorporated by reference herein. This is particularly important where the sample may contain components that can interfere with the amplification reaction, such as, for example, heme present in a blood sample. *See* Ryder *et al.*, "Amplification of Nucleic Acids from Mononuclear Cells Using Iron Complexing and Other Agents," U.S. Patent No. 5,639,599, the contents of which are hereby incorporated by reference herein. Methods to prepare target nucleic acids from various sources for amplification are well known to those of ordinary skill in the art. Target nucleic acids of the present invention may be purified to some degree prior to the amplification reactions described herein, but in other cases, the sample is added to the amplification reaction without any further manipulations.

The term "target sequence" refers to the particular nucleotide sequence of the target nucleic acid which is to be amplified. The "target sequence" includes the complexing sequences to which oligonucleotides (*e.g.*, priming oligonucleotides and/or promoter oligonucleotides) complex during the processes of the present invention. Where the target nucleic acid is originally single-stranded, the term "target sequence" will also refer to the sequence complementary to the "target sequence" as present in the target nucleic acid. Where the "target nucleic acid" is originally double-stranded, the term "target sequence" refers to both the sense (+) and antisense (-) strands. In choosing a target sequence, the skilled artisan will understand that a "unique" sequence should be chosen so as to distinguish between unrelated or closely related target nucleic acids. As will be understood by those of ordinary skill in the art, "unique" sequences are judged from the testing environment. At least the sequences recognized by the detection probe (as described in more detail elsewhere herein) should be unique in the environment being tested, but need not be unique within the universe of all possible sequences. Furthermore, even though the target sequence should contain a "unique" sequence for recognition by a detection probe, it is not always the case that the priming oligonucleotide and/or promoter oligonucleotide are recognizing "unique" sequences. In some embodiments, it may be desirable to choose a target sequence which is common to a family of related organisms, for example, a sequence which is common to all HIV strains that might be in a sample. In other situations, a very highly specific target sequence, or a target sequence having at least a highly specific region recognized by the detection probe, would be chosen so as to distinguish between closely related organisms, for example, between pathogenic and non-pathogenic *E. coli*. A target sequence of the present invention may be of any practical length. A minimal target sequence includes the region which hybridizes to the priming oligonucleotide (or the complement thereof), the region which hybridizes to the

hybridizing region of the promoter oligonucleotide (or the complement thereof), and a region used for detection, *e.g.*, a region which hybridizes to a detection probe, described in more detail elsewhere herein. The region which hybridizes to the detection probe may overlap with or be contained within the region which hybridizes with the priming oligonucleotide (or its complement) or the hybridizing region of the promoter oligonucleotide (or its complement). In addition to the minimal requirements, the optimal length of a target sequence depends on a number of considerations, for example, the amount of secondary structure, or self-hybridizing regions in the sequence. Determining the optimal length is easily accomplished by those of ordinary skill in the art using routine optimization methods. Typically, target sequences of the present invention range from about 100 nucleotides in length to from about 150 to about 250 nucleotides in length. The optimal or preferred length may vary under different conditions, which can easily be tested by one of ordinary skill in the art according to the methods described herein. The term "amplicon" refers to a nucleic acid molecule that is generated during an amplification procedure and is substantially complementary or identical to a sequence contained within the target sequence.

17. Template

A "template" is a nucleic acid molecule that is being copied by a nucleic acid polymerase. A template may be single-stranded, double-stranded or partially double-stranded, depending on the polymerase. The synthesized copy is complementary to the template or to at least one strand of a double-stranded or partially double-stranded template. Both RNA and DNA are typically synthesized in the 5'-to-3' direction and the two strands of a nucleic acid duplex are aligned so that the 5'-termini of the two strands are at opposite ends of the duplex (and, by necessity, so then are the 3'-termini). While according to the present invention, a "target sequence" is always a "template," templates can also include secondary primer extension products and amplification products.

18. DNA-dependent DNA Polymerase

A "DNA-dependent DNA polymerase" is an enzyme that synthesizes a complementary DNA copy from a DNA template. Examples are DNA polymerase I from *E. coli*, bacteriophage T7 DNA polymerase, or DNA polymerases from bacteriophages T4, Phi-29, M2, or T5. DNA-dependent DNA polymerases of the present invention may be the naturally occurring enzymes isolated from bacteria or bacteriophages or expressed recombinantly, or may be modified or "evolved" forms which have been engineered to possess certain desirable characteristics, *e.g.*, thermostability, or the ability to recognize or synthesize a DNA strand from various modified templates. All known DNA-dependent DNA

polymerases require a complementary primer to initiate synthesis. It is known that under suitable conditions a DNA-dependent DNA polymerase may synthesize a complementary DNA copy from an RNA template. RNA-dependent DNA polymerases (described below) typically also have DNA-dependent DNA polymerase activity.

19. DNA-dependent RNA Polymerase (Transcriptase)

A "DNA-dependent RNA polymerase" or "transcriptase" is an enzyme that synthesizes multiple RNA copies from a double-stranded or partially-double-stranded DNA molecule having a promoter sequence that is usually double-stranded. The RNA molecules ("transcripts") are synthesized in the 5'-to-3' direction beginning at a specific position just downstream of the promoter. Examples of transcriptases are the DNA-dependent RNA polymerase from *E. coli* and bacteriophages T7, T3, and SP6.

20. RNA-dependent DNA polymerase (Reverse Transcriptase)

An "RNA-dependent DNA polymerase" or "reverse transcriptase" ("RT") is an enzyme that synthesizes a complementary DNA copy from an RNA template. All known reverse transcriptases also have the ability to make a complementary DNA copy from a DNA template; thus, they are both RNA- and DNA-dependent DNA polymerases. RTs may also have an RNAse H activity. Preferred is reverse transcriptase derived from Maloney murine leukemia virus (MMLV-RT). A primer is required to initiate synthesis with both RNA and DNA templates.

21. Selective RNAses

As used herein, a "selective RNAse" is an enzyme that degrades the RNA portion of an RNA:DNA duplex but not single-stranded RNA, double-stranded RNA or DNA. An exemplary selective RNAse is RNAse H. Enzymes other than RNAse H which possess the same or similar activity are also contemplated in the present invention. Selective RNAses may be endonucleases or exonucleases. Most reverse transcriptase enzymes contain an RNAse H activity in addition to their polymerase activities. However, other sources of the RNAse H are available without an associated polymerase activity. The degradation may result in separation of RNA from a RNA:DNA complex. Alternatively, a selective RNAse may simply cut the RNA at various locations such that portions of the RNA melt off or permit enzymes to unwind portions of the RNA. Other enzymes which selectively degrade RNA target sequences or RNA products of the present invention will be readily apparent to those of ordinary skill in the art.

22. Sense/Antisense Strand(s)

Discussions of nucleic acid synthesis are greatly simplified and clarified by adopting terms to name the two complementary strands of a nucleic acid duplex. Traditionally, the strand encoding the sequences used to produce proteins or structural RNAs are designated as the "sense (+)" strand and its complement the "antisense (-)" strand. It is now known that in many cases, both strands are functional, and the assignment of the designation "sense" to one and "antisense" to the other must then be arbitrary. Nevertheless, the terms are very useful for designating the sequence orientation of nucleic acids and will be employed herein for that purpose.

23. Specificity of the System

The term "specificity," in the context of an amplification system, is used herein to refer to the characteristic of an amplification system which describes its ability to distinguish between target and non-target sequences dependent on sequence and assay conditions. In terms of a nucleic acid amplification, specificity generally refers to the ratio of the number of specific amplicons produced to the number of side-products (*i.e.*, the signal-to-noise ratio), described in more detail below.

24. Sensitivity

The term "sensitivity" is used herein to refer to the precision with which a nucleic acid amplification reaction can be detected or quantitated. The sensitivity of an amplification reaction is generally a measure of the smallest copy number of the target nucleic acid that can be reliably detected in the amplification system, and will depend, for example, on the detection assay being employed, and the specificity of the amplification reaction, *i.e.*, the ratio of specific amplicons to side-products.

25. Amplification Conditions

By "amplification conditions" is meant conditions permitting nucleic acid amplification according to the present invention. Amplification conditions may, in some embodiments, be less stringent than "stringent hybridization conditions" as described herein. Oligonucleotides used in the amplification reactions of the present invention hybridize to their intended targets under amplification conditions, but may or may not hybridize under stringent hybridization conditions. On the other hand, detection probes of the present invention hybridize under stringent hybridization conditions. While the Examples section *infra* provides preferred amplification conditions for amplifying target nucleic acid sequences according to the present invention, other acceptable conditions to carry out nucleic acid

amplifications according to the present invention could be easily ascertained by someone having ordinary skill in the art depending on the particular method of amplification employed.

* * * * *

The present invention provides an autocatalytic amplification method which synthesizes large numbers of RNA copies of an RNA or DNA target sequence with high specificity and sensitivity. An important aspect of the present invention is the minimal production of side-products during the amplification. Examples of side-products include oligonucleotide dimers and self-replicating molecules. The target nucleic acid contains the target sequence to be amplified. The target sequence is that region of the target nucleic acid which is defined on either end by priming oligonucleotides, promoter oligonucleotides, and, optionally, a binding molecule, *e.g.*, a terminating oligonucleotide or a modifying oligonucleotide (described in more detail below), and/or the natural target nucleic acid termini, and includes both the sense and antisense strands. Promoter oligonucleotides of the present invention are modified to prevent the synthesis of DNA therefrom. Preferably, the promoter oligonucleotides comprise a blocking moiety attached at their 3'-termini to prevent primer extension in the presence of a polymerase. Indeed, according to the present invention, at least about 80% of the oligonucleotides present in the amplification reaction which comprise a promoter further comprise a 3'-blocking moiety. In further embodiments, at least about 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% of the oligonucleotides provided to the amplification reaction which comprise a promoter are further modified to comprise a 3'-blocking moiety. In a specific embodiment, any oligonucleotide used in an amplification reaction of the present invention which comprises a promoter sequence must further comprise a 3'-terminus blocking moiety.

One embodiment of the present invention comprises amplification of a target nucleic acid comprising an RNA target sequence. *See* Figures 1A and 1B. The target nucleic acid has indeterminate 3'- and 5'-ends relative to the desired RNA target sequence. The target nucleic acid is treated with a priming oligonucleotide which has a base region sufficiently complementary to a 3'-region of the RNA target sequence to hybridize therewith. *See* Step 1 of Figures 1A and 1B. Priming oligonucleotides are designed to hybridize to a suitable region of any desired target sequence, according to primer design methods well known to those of ordinary skill in the art. Suitable priming oligonucleotides are described in more detail herein. While a priming oligonucleotide of the present invention can optionally include a non-hybridizing base region situated 5' to the region which hybridizes with the target sequence, according to the present invention the 5' region of a priming oligonucleotide does not include a promoter sequence recognized by an RNA polymerase. Additionally, the 5'-end

of the priming oligonucleotide may include one or modifications which improve the binding properties (*e.g.*, hybridization or base stacking) of the priming oligonucleotide to a target sequence or RNA amplification product, as discussed more fully *infra*, provided the modifications do not substantially interfere with the priming function of the priming oligonucleotide or cleavage of a template RNA to which the priming oligonucleotide is hybridized. In the presence of nucleoside triphosphates and buffers, salts and cofactors, the 3'-end of the priming oligonucleotide is extended by an appropriate DNA polymerase, *e.g.*, an RNA- dependent DNA polymerase ("reverse transcriptase") in an extension reaction using the RNA target sequence as a template to give a DNA primer extension product which is complementary to the RNA template. *See* Steps 2 and 3 of Figures 1A and 1B.

The DNA primer extension product is separated (at least partially) from the RNA template using an enzyme which degrades the RNA template. *See* Step 4 of Figures 1A and 1B. Suitable enzymes, *i.e.*, "selective RNAses," are those which act on the RNA strand of an RNA:DNA complex, and include enzymes which comprise an RNAse H activity. Some reverse transcriptases include an RNAse H activity, including those derived from Moloney murine leukemia virus ("MMLV") and avian myeloblastosis virus ("AMV"). According to this method, the selective RNAse may be provided as an RNAse H activity of a reverse transcriptase, or may be provided as a separate enzyme, *e.g.*, as an *E. coli* RNAse H or a *T. thermophilus* RNAse H. Other enzymes which selectively degrade RNA present in an RNA:DNA duplex may also be used.

In certain specific embodiments, the method of the present invention further comprises treating the target nucleic acid as described above to limit the length of the DNA primer extension product to a certain desired length. Such length limitation is typically carried out through use of a "binding molecule" which hybridizes to or otherwise binds to the RNA target nucleic acid adjacent to or near the 5'-end of the desired target sequence. *See* Step 1 of FIG. 1A. In certain embodiments, a binding molecule comprises a base region. The base region may be DNA, RNA, a DNA:RNA chimeric molecule, or an analog thereof. Binding molecules comprising a base region may be modified in one or more ways, as described elsewhere herein. Suitable binding molecules include, but are not limited to, a binding molecule comprising a terminating oligonucleotide or a terminating protein that binds RNA and prevents primer extension past its binding region, or a binding molecule comprising a modifying molecule, for example, a modifying oligonucleotide such as a "digestion" oligonucleotide that directs hydrolysis of that portion of the RNA target hybridized to the digestion oligonucleotide, or a sequence-specific nuclease that cuts the RNA target.

A terminating oligonucleotide of the present invention has a 5'-base region sufficiently complementary to the target nucleic acid at a region adjacent to, near to, or overlapping with

the 5'-end of the target sequence, to hybridize therewith. In certain embodiments, a terminating oligonucleotide is synthesized to include one or more modified nucleotides. For example, certain terminating oligonucleotides of the present invention comprise one or more 2'-O-methyl ribonucleotides, or are synthesized entirely of 2'-O-methyl ribonucleotides. *See, e.g.*, Majlessi *et al.* (1998) *Nucleic Acids Res.*, 26, 2224-2229. A terminating oligonucleotide of the present invention typically also comprises a blocking moiety at its 3'-end to prevent the terminating oligonucleotide from functioning as a primer for a DNA polymerase. In some embodiments, the 5'-end of a terminating oligonucleotide of the present invention overlaps with and is complementary to at least about 2 nucleotides of the 5'-end of the target sequence. Typically, the 5'-end of a terminating oligonucleotide of the present invention overlaps with and is complementary to at least 3, 4, 5, 6, 7, or 8 nucleotides of the 5'-end of the target sequence, but no more than about 10 nucleotides of the 5'-end of the target sequence. (As used herein, the term "end" refers to a 5'- or 3'-region of an oligonucleotide, nucleic acid or nucleic acid region which includes, respectively, the 5'- or 3'-terminal base of the oligonucleotide, nucleic acid or nucleic acid region.) Suitable terminating oligonucleotides are described in more detail herein.

To the extent that a terminating oligonucleotide has a 5' base region which overlaps with the target sequence, it may be desirable to introduce one or more base mismatches into the 5'-end of the first region of a promoter oligonucleotide in order to minimize or prevent hybridization of the terminating oligonucleotide to the promoter oligonucleotide, as the formation of terminating oligonucleotide:promoter oligonucleotide hybrids may negatively affect the rate of an amplification reaction. While one base mismatch in the region of overlap generally should be sufficient, the exact number needed will depend upon factors such as the length and base composition of the overlapping region, as well as the conditions of the amplification reaction. Despite the possible benefits of a modified promoter oligonucleotide, it should be noted that mutations in the first region of the promoter oligonucleotide could render it a poorer template for amplification. Moreover, it is entirely possible that in a given amplification system the formation of terminating oligonucleotide:promoter oligonucleotide hybrids advantageously prevents or interferes with the formation of priming oligonucleotide:promoter oligonucleotides hybrids with a 3'-end available for primer extension. *See* FIG. 5 (formation of primer-dependent side-products).

A modifying oligonucleotide provides a mechanism by which the 3'-terminus of the primer extension product is determined. A modifying oligonucleotide may provide a motif comprising one or more bases in the vicinity of the 5'-end of the RNA target sequence which facilitates termination of primer extension by means of a modifying enzyme, *e.g.*, a nuclease.

Alternatively, a modifying oligonucleotide might be tethered to a specific modifying enzyme or to a chemical which can then terminate primer extension.

One specific modifying oligonucleotide is a digestion oligonucleotide. A digestion oligonucleotide is comprised of DNA, preferably a stretch of at least about 6 deoxyribonucleotides. The digestion oligonucleotide hybridizes to the RNA template and the RNA of the RNA:DNA hybrid is digested by a selective RNAse as described herein, *e.g.*, by an RNAse H activity.

The single-stranded DNA primer extension product, or "first" DNA primer extension product, which has either a defined 3'-end or an indeterminate 3'-end, is then treated with a promoter oligonucleotide which comprises a first region sufficiently complementary to a 3'-region of the DNA primer extension product to hybridize therewith, a second region comprising a promoter for an RNA polymerase, *e.g.*, T7 polymerase, which is situated 5' to the first region, *e.g.*, immediately 5' to or spaced from the first region, and modified to prevent the promoter oligonucleotide from functioning as a primer for a DNA polymerase (*e.g.*, the promoter oligonucleotide includes a blocking moiety attached at its 3'-terminus). *See* Step 5 of Figures 1A and 1B. Upon identifying a desired hybridizing "first region," suitable promoter oligonucleotides can be constructed by one of ordinary skill in the art using only routine procedures. Those of ordinary skill in the art will readily understand that a promoter region has certain nucleotides which are required for recognition by a given RNA polymerase. In addition, certain nucleotide variations in a promoter sequence might improve the functioning of the promoter with a given enzyme, including the use of insertion sequences.

Insertion sequences are positioned between the first and second regions of promoter oligonucleotides and function to increase amplification rates. The improved amplification rates may be attributable to several factors. First, because an insertion sequence increases the distance between the 3'-end and the promoter sequence of a promoter oligonucleotide, it is less likely that a polymerase, *e.g.*, reverse transcriptase, bound at the 3'-end of the promoter oligonucleotide will interfere with binding of the RNA polymerase to the promoter sequence, thereby increasing the rate at which transcription can be initiated. Second, the insertion sequence selected may itself improve the transcription rate by functioning as a better template for transcription than the target sequence. Third, since the RNA polymerase will initiate transcription at the insertion sequence, the primer extension product synthesized by the priming oligonucleotide, using the RNA transcription product as a template, will contain the complement of the insertion sequence toward the 3'-end of the primer extension product. By providing a larger target binding region, *i.e.*, one which includes the complement of the insertion sequence, the promoter oligonucleotide may bind to the primer extension product faster, thereby leading to the production of additional RNA transcription products sooner.

Insertion sequences are preferably 5 to 20 nucleotides in length and should be designed to minimize intramolecular folding and intermolecular binding with other oligonucleotides present in the amplification reaction mixture. Programs which aid in minimizing secondary structure are well known in the art and include Michael Zucker's *mfold* software for predicting RNA and DNA secondary structure using nearest neighbor thermodynamic rules. The latest version of Michael Zucker's *mfold* software can be accessed on the Web at www.bioinfo.rpi.edu/applications/mfold using a hypertext transfer protocol (http) in the URL. Currently preferred insertion sequences include the nucleotide sequences of SEQ ID Nos. 1 and 2 in combination with the T7 RNA polymerase promoter sequence of SEQ ID NO:3. *See* Ikeda *et al.* (1992) *J. Biol. Chem.* 267, 2640-2649. Other useful insertion sequences may be identified using *in vitro* selection methods well known in the art.

Assaying promoter oligonucleotides with variations in the promoter sequences is easily carried out by the skilled artisan using routine methods. Furthermore, if it is desired to utilize a different RNA polymerase, the promoter sequence in the promoter oligonucleotide is easily substituted by a different promoter. Substituting different promoter sequences is well within the understanding and capabilities of those of ordinary skill in the art. It is important to note that according to the present invention, promoter oligonucleotides provided to the amplification reaction mixture are modified to prevent the initiation of DNA synthesis from their 3'-termini, and preferably comprise a blocking moiety attached at their 3'-termini. Furthermore, terminating oligonucleotides and capping oligonucleotides, and even probes used in the methods of the present invention also optionally comprise a blocking moiety attached at their 3'-termini.

Where a terminating oligonucleotide is used, the first region of the promoter oligonucleotide is designed to hybridize with a desired 3'-end of the DNA primer extension product with substantial, but not necessarily exact, precision. Subsequently, the second region of the promoter oligonucleotide may act as a template, allowing the first DNA primer extension product to be further extended to add a base region complementary to the second region of the promoter oligonucleotide, *i.e.*, the region comprising the promoter sequence, rendering the promoter double-stranded. *See* Steps 6 and 7 of FIG. 1A. An RNA polymerase which recognizes the promoter binds to the promoter sequence, and initiates transcription of multiple RNA copies complementary to the DNA primer extension product, which copies are substantially identical to the target sequence. By "substantially identical" it is meant that the multiple RNA copies may have additional nucleotides either 5' or 3' relative to the target sequence, or may have fewer nucleotides either 5' or 3' relative to the target sequence, depending on, *e.g.*, the boundaries of "the target sequence," the transcription initiation point, or whether the priming oligonucleotide comprises additional nucleotides 5' of the primer

region (*e.g.*, a linked "cap" as described herein). Where a target sequence is DNA, the sequence of the RNA copies is described herein as being "substantially identical" to the target sequence. It is to be understood, however, that an RNA sequence which has uridine residues in place of the thymidine residues of the DNA target sequence still has a "substantially identical" sequence. The RNA transcripts so produced may automatically recycle in the above system without further manipulation. Thus, this reaction is autocatalytic. In those embodiments where a binding molecule or other means for terminating a primer extension reaction is not used, the first region of the promoter oligonucleotide is designed to hybridize with a selected region of the first DNA primer extension product which is expected to be 5' to the 3'-terminus of the first DNA primer extension product, but since the 3'-terminus of the first DNA primer extension product is indeterminate, the region where the promoter oligonucleotide hybridizes probably will not be at the actual 3'-end of the first DNA primer extension product. According to this embodiment, it is generally the case that at least the 3'-terminal base of the first DNA primer extension product does not hybridize to the promoter oligonucleotide. *See* Step 5 of FIG. 1B. Thus, according to this embodiment the first DNA primer extension product will likely not be further extended to form a double-stranded promoter.

Surprisingly, the inventors discovered that the formation of a double-stranded promoter sequence through extension of a template nucleic acid is not necessary to permit initiation of transcription of RNA complementary to the first DNA primer extension product. *See* Step 6 of FIG. 1B. The resulting "first" RNA products are substantially identical to the target sequence, having a 5'-end defined by the transcription initiation point, and a 3'-end defined by the 5'-end of the first DNA primer extension product. *See* Step 7 of FIG. 1B. As illustrated in FIG. 1B, a sufficient number of first RNA products are produced to automatically recycle in the system without further manipulation. The priming oligonucleotide hybridizes to the 3'-end of the first RNA products, and is extended by a DNA polymerase to form a second DNA primer extension product. Unlike the first DNA primer extension product formed without the use of a terminating oligonucleotide or other binding molecule, the second DNA primer extension product has a defined 3'-end which is complementary to the 5'-ends of the first RNA products. *See* Steps 8-10 of FIG. 1B. The second DNA primer extension product is separated (at least partially) from the RNA template using an enzyme which selectively degrades the RNA template. *See* Step 11 of FIG. 1B. The single-stranded second DNA primer extension product is then treated with a promoter oligonucleotide as described above, and the second region of the promoter oligonucleotide acts as a template, allowing the second DNA primer extension product to be further extended to add a base region complementary to the second region of the promoter oligonucleotide, *i.e.*,

the region comprising the promoter sequence, rendering the promoter double-stranded. *See* Steps 12-14 of FIG. 1B. An RNA polymerase which recognizes the promoter binds to the promoter sequence, and initiates transcription of multiple "second" RNA products complementary to the second DNA primer extension product, and substantially identical to the target sequence. *See* Step 15 of FIG. 1B. The second RNA transcripts so produced automatically recycle in the above system without further manipulation. Thus, this reaction is autocatalytic. *See* Steps 7-15 of FIG. 1B.

In another embodiment, the present invention is drawn to a method of synthesizing multiple copies of a target sequence from a target nucleic acid comprising a DNA target sequence. This embodiment is diagramed in FIG. 1C. The target nucleic acid may be either single-stranded, partially single-stranded, or double-stranded DNA. When the DNA is double-stranded, it is denatured, or partially denatured, prior to amplification. The DNA target nucleic acid need not have a defined 3'-end. The single-stranded, partially single-stranded, or denatured DNA target nucleic acid is treated with a promoter oligonucleotide as described above. The first region of the promoter oligonucleotide is designed to hybridize with a selected region of the target nucleic acid in the 3'-region of the desired target sequence, but since the 3'-end of the target nucleic acid need not be coterminal with the 3'-end of the target sequence, the region where the promoter oligonucleotide hybridizes will likely not be at or near the 3'-end of the target nucleic acid sequence. *See* Step 1 of FIG. 1C. Thus, the promoter region of the promoter oligonucleotide will likely remain single-stranded.

As noted above, the inventors surprisingly discovered that it is not necessary for the single-stranded promoter sequence on the promoter oligonucleotide to form a double-stranded promoter through extension of a template nucleic acid in order for the promoter sequence to be recognized by the corresponding RNA polymerase and, in this case, initiate transcription of RNA complementary to the DNA target sequence. *See* Step 2 of FIG. 1C. The resulting "first RNA products" have a 5'-end defined by the transcription initiation point for the promoter, however, the 3'-region will remain indeterminate. *See* Step 3 of FIG. 1C. These first RNA products are then treated with a priming oligonucleotide. The priming oligonucleotide hybridizes to a region of the first RNA products at a position complementary to a 5' region of the desired target sequence, and is extended by a DNA polymerase to form a DNA primer extension product. *See* Steps 4-6 of FIG. 1C. This DNA primer extension product has a 5'-end coinciding with the 5'-end of the priming oligonucleotide, and a 3'-end coinciding with the 5'-end of the first RNA products. *See* Step 6 of FIG. 1C. The DNA primer extension product is separated (at least partially) from its RNA template using an enzyme which selectively degrades the RNA template, as described above. *See* Step 7 of FIG. 1C. The DNA primer extension product is then treated with the promoter oligonucleotide, as

described above, and the second region of the promoter oligonucleotide acts as a template, allowing the DNA primer extension product to be further extended to add a base region complementary to the second region of the promoter oligonucleotide, *i.e.*, the region comprising the promoter, rendering the promoter double-stranded. *See* Steps 8-10 of FIG. 1C. An RNA polymerase which recognizes the promoter binds to the promoter sequence, and initiates transcription of multiple RNA products complementary to the DNA primer extension product. *See* Step 11 of FIG. 1C. The sequence of these "second RNA products" is substantially complementary to the desired target sequence. The RNA products so produced automatically recycle in the above system without further manipulation. Thus, this reaction is autocatalytic. *See* Steps 3-11 of FIG. 1C.

In yet another embodiment, the present invention relates to a method of synthesizing multiple copies of a target sequence from a target nucleic acid comprising a DNA target sequence. One aspect of this embodiment is illustrated in FIG. 1D. Prior to amplification, the target nucleic acid in this embodiment may be single-stranded, partially single-stranded, or double-stranded. If a region of the target nucleic acid containing the target sequence is double-stranded, then the target nucleic acid may be denatured, or partially denatured, prior to initiating amplification to render the target sequence accessible to the priming oligonucleotide. *See* Step 1 of FIG. 1D. Denaturation may be effected by heat, high pH, and/or low ionic strength. (Useful chemical denaturants include imidazole, dimethyl sulfoxide, formamide, urea and/or sodium hydryoxide.) The accessible target sequence is treated with a priming oligonucleotide, as described *infra*, which hybridizes to a 3'-end of the target sequence. *See* Step 2 of FIG. 1D. In the presence of nucleoside triphosphates and buffers, salts and cofactors, the 3'-end of the priming oligonucleotide is extended by a DNA polymerase in an extension reaction using the DNA target sequence as a template to give a first DNA primer extension product that is complementary to the template. *See* Steps 3 and 4 of FIG. 1D. In one aspect of this embodiment, a binding molecule is included which limits the length of the first DNA primer extension product by hybridizing or otherwise binding to the target nucleic acid adjacent to or near the 5'-end of the target sequence. The binding molecule may be any of the binding molecules described herein that is appropriate for a target nucleic acid containing a DNA target sequence, but is preferably a terminating oligonucleotide. *See* Steps 2 to 4 of FIG. 1D. One advantage of a terminating oligonucleotide is that it does not require the use of a restriction endonuclease to create a defined 3'-end on the first DNA primer extension product.

To separate the first DNA primer extension product from the template, the target nucleic acid can be treated with a displacer oligonucleotide in another aspect of this embodiment. *See* Step 5 of FIG. 1D. The displacer oligonucleotide has a priming function

and is designed to hybridize to the target nucleic acid upstream from the priming oligonucleotide (referred to as the "forward priming oligonucleotide" in this embodiment). By "upstream" is meant that a 3'-end of the displacer oligonucleotide hybridizes to the target nucleic acid upstream from a 3'-end of the forward priming oligonucleotide. Thus, the displacer oligonucleotide and the forward priming oligonucleotide may hybridize to overlapping or distinct regions of the target nucleic acid. In preferred embodiments, the 3'-terminus of the displacer oligonucleotide is adjacent to or spaced up to 5 to 35 bases from the 5'-terminus of the forward priming oligonucleotide relative to the target nucleic acid (*i.e.*, the target nucleic acid has up to 5 to 35, contiguous unbound nucleotides situated between the 3'-terminal base of the displacer oligonucleotide and the 5'-terminal base of the priming oligonucleotide when both oligonucleotides are hybridized to the target nucleic acid). The displacer oligonucleotide is generally from 10 to 50 nucleotides in length and may include one or more modifications at the 5'-end which improve the binding properties (*e.g.*, hybridization or base stacking) of the displacer oligonucleotide to the target nucleic acid, provided that the modifications do not substantially interfere with the priming function of the displacer oligonucleotide. The displacer oligonucleotide and the forward priming oligonucleotide are designed to hybridize to the target nucleic acid under the same conditions. The target nucleic acid is preferably treated with the displacer oligonucleotide after the forward priming oligonucleotide has had sufficient time to hybridize to the target nucleic acid. Alternatively, the target nucleic acid is treated with both the displacer oligonucleotide and the forward priming oligonucleotide before exposing the mixture to a polymerase suitable for extending the 3'-ends of the displacer oligonucleotide and the forward priming oligonucleotide. In the presence of the DNA polymerase, the 3'-end of the displacer oligonucleotide is extended in a template-dependent manner to form a second DNA primer extension product which displaces the first DNA primer extension product from the target nucleic acid, thereby making it available for hybridization to a promoter oligonucleotide. *See* Steps 6 and 7 of FIG. 1D. In an alternative approach, conditions could be established whereby the promoter oligonucleotide gains access the first DNA primer extension product through stand invasion facilitated by, for example, DNA breathing (*e.g.*, AT rich regions), low salt conditions, and/or the use of DMSO and/or osmolytes, such as betaine. The promoter oligonucleotide of this embodiment is the same as that described above and, likewise, is modified to prevent the promoter oligonucleotide from functioning as a priming oligonucleotide for a DNA polymerase (*e.g.*, the promoter oligonucleotide includes a blocking moiety at its 3'-terminus).

Where a terminating oligonucleotide or other binding molecule capable of determining the 3'-end of the first DNA primer extension product is used, a first region of the promoter oligonucleotide is designed to hybridize to the 3'-end of the first DNA primer extension

product with sufficient precision that a second region of the promoter oligonucleotide acts as a template, allowing the first DNA primer extension product to be further extended to add a base region complementary to the second region of the promoter oligonucleotide. *See* Steps 8-10 of FIG. 1D. This extension reaction renders a region containing the promoter sequence double-stranded. An RNA polymerase which recognizes the promoter then binds to the promoter sequence and initiates transcription of multiple RNA copies complementary to the first DNA primer extension product, where the RNA copies are substantially identical to the target sequence, as this term is defined hereinabove. *See* Steps 11 and 12 of FIG. 1D. The RNA transcripts so produced may automatically recycle in the method of this embodiment and, thus, the reaction is autocatalytic. *See* Steps 9-17 of FIG.1D. In the subsequent rounds of amplification, a third DNA primer extension product comprising the promoter oligonucleotide and complementary to the RNA transcript is formed, and this DNA primer extension product is then separated from the RNA transcript (at least partially) using an enzyme which selectively degrades the RNA transcript (*e.g.*, an enzyme having an RNase H activity, such as a reverse transcriptase derived from MMLV or AMV). *See* Steps 13-16 of FIG. 1D.

In those aspects of this embodiment where a binding molecule or other means for terminating a primer extension reaction are not used, the first region of the promoter oligonucleotide is designed to hybridize to a selected region of the first DNA primer extension product which is expected to be 5' to the 3'-terminus of the first DNA primer extension product. Since the 3'-terminus of the first DNA primer extension product is indeterminate, the region where the promoter oligonucleotide hybridizes probably will not be at the actual 3'-end of the first DNA primer extension product. According to this embodiment, it is generally the case that at least the 3'-terminal base of the first DNA primer extension product does not hybridize to the promoter oligonucleotide. Thus, according to this embodiment the first DNA primer extension product will likely not be further extended to form a double-stranded promoter. As discussed above, the inventors surprisingly discovered that the formation of a double-stranded promoter sequence through extension of a template nucleic acid is not necessary to permit initiation of transcription of RNA complementary to the first DNA primer extension product. In subsequent rounds of amplification, however, the RNA transcripts will include a defined 3'-end and DNA primer extension products complementary to the RNA transcripts will hybridize to the promoter oligonucleotide and be extended to add a sequence complementary to a region of the promoter oligonucleotide containing the promoter sequence.

The inventors also discovered that the rate of amplification could be enhanced by providing an extender oligonucleotide to a reaction mixture, as diagramed in Figures 2A-2D.

(Step 5 of Figures 2A and 2B, Step 4 of FIG. 2C, and Step 8 of FIG. 2D.) An extender oligonucleotide is generally 10 to 50 nucleotides in length and hybridizes to a DNA template (*i.e.*, the DNA target sequence or any of the DNA primer extension products described herein) downstream from a promoter oligonucleotide. When included, the 5'-terminal base of the extender oligonucleotide is positioned near or adjacent to the 3'-terminal base of the promoter oligonucleotide when both oligonucleotides are hybridized to a DNA template. (By "adjacent to" is meant that the DNA template has no unbound bases situated between the 3'-terminal base of the promoter oligonucleotide and the 5'-terminal base of the extender oligonucleotide when both oligonucleotides are hybridized to the DNA template.) Most preferably, the extender oligonucleotide hybridizes to a DNA template such that the 5'-terminal base of the extender oligonucleotide is spaced no more than three nucleotides from the 3'-terminal base of the promoter oligonucleotide relative to the DNA template (*i.e.*, the DNA template has a maximum of three, contiguous unbound nucleotides situated between the 3'-terminal base of the promoter oligonucleotide and the 5'-terminal base of the extender oligonucleotide when both oligonucleotides are hybridized to the DNA template). To prevent the extender oligonucleotide from functioning as a priming oligonucleotide in a primer extension reaction, the extender oligonucleotide preferably includes a 3'-terminal blocking moiety. While not wishing to be bound by theory, it is believed that the phosphate at the 3'-end of the extender oligonucleotide functions to draw the DNA-dependent DNA polymerase (*e.g.*, reverse transcriptase) farther away from the promoter sequence of the promoter oligonucleotide, thereby minimizing interference with the binding and progress of the RNA polymerase in transcription. It is also possible that the extender oligonucleotide facilitates faster trancription reactions by limiting secondary structure within the target sequence.

In one aspect, the present invention relates to minimizing side-product formation in nucleic acid amplification reactions. One type of side-product is referred to herein as an "oligonucleotide dimer." This side-product occurs when a priming oligonucleotide base-pairs non-specifically with another nucleic acid in the amplification reaction, *e.g.*, the promoter oligonucleotide. Since the priming oligonucleotide can be extended via a DNA polymerase, a double-stranded form of the promoter oligonucleotide can result, which can be transcribed into non-specific, amplifiable side-products. To prevent priming oligonucleotides from participating in the formation of oligonucleotide dimers, one option is to add a short complementary nucleotide "cap" to the 3'-end of the priming oligonucleotide. *See* Figures 6A and 6B. A cap is thought to reduce non-specific hybridization between the priming oligonucleotide and other nucleic acids in the reaction, *e.g.*, the promoter oligonucleotide, thereby eliminating or substantially reducing the production of "oligonucleotide-dimer" side-products as compared to amplification reactions carried out under identical conditions, but

without the use of a cap. As used herein, a cap comprises a base region complementary to a region at the 3'-end of the priming oligonucleotide which is preferably pre-hybridized to the priming oligonucleotide prior to its introduction into an amplification reaction mixture. A suitable cap length will vary based on base content, stringency conditions, etc., but will typically hybridize to up to 3, 6, 9, 12, 15, 18, or 20 contiguous or non-contiguous nucleotides at the '3-end of the priming oligonucleotide. Suitable caps preferably range from 5 to 10 bases in length. The length of the complementary cap region is dependent on several variables, for example, the melting temperature of the double-stranded hybrid formed with the 3'-end of the priming oligonucleotide. In general, an efficient cap will specifically hybridize to a region at the 3'-end of the priming oligonucleotide more strongly than any non-specific reactions with other oligonucleotides present in the amplification reaction, but will be readily displaced in favor of specific hybridization of the priming oligonucleotide with the desired template. Exemplary caps comprise, or alternately consist essentially of, or alternately consist of an oligonucleotide from 5 to 7 bases in length which hybridizes to a region at the 3'-end of the priming oligonucleotide, such that the 5'-terminal base of the cap hybridizes to the 3'-terminal base of the priming oligonucleotide. Typically, a cap will hybridize to no more than 8, 9, or 10 nucleotides of a region at the 3'-end of the priming oligonucleotide.

A cap may take the form of a capping oligonucleotide or a base region attached to the 5'-end of the priming oligonucleotide, either directly or through a linker. *See* Figures 6A and 6B. A capping oligonucleotide is synthesized as a separate oligonucleotide from the priming oligonucleotide, and normally comprises a blocking moiety at its 3'-terminus to prevent primer extension by a DNA polymerase, as illustrated in FIG. 6A. Alternatively, the cap comprises a base region complementary to a region at the 3'-end of the priming oligonucleotide, which is connected to the 5'-end of the priming oligonucleotide via a linking region comprising, alternately consisting essentially of, or alternately consisting of 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. *See* FIG. 6B. Typically, the nucleotides in the linking region are abasic nucleotides. By "abasic nucleotide" is meant a nucleotide comprising a phosphate group and a sugar group, but not a base group. Constructing a priming oligonucleotide with a cap attached to its 5'-end simplifies oligonucleotide synthesis by requiring the synthesis of only a single oligonucleotide comprising both the priming portion and the cap.

The caps described herein may also be used with other oligonucleotides provided to a reaction mixture, including displacer and extender oligonucleotides, which have free 3'-hydroxyls that are capable of participating in an extension reaction in the presence of a DNA polymerase. The importance of caps increases with the number of oligonucleotides having free 3'-hydroxyls.

In any of the embodiments described above, once a desired region for the target sequence is identified, that region can be analyzed to determine where selective RNAse degradation will optimally cause cuts or removal of sections of RNA from the RNA:DNA duplex. Analyses can be conducted to determine the effect of the RNAse degradation of the target sequence by RNAse H activity present in AMV reverse transcriptase or MMLV reverse transcriptase, by an exogenously added selective enzyme with an RNAse activity, *e.g.*, *E. coli* RNAse H, or selective enzymes with an RNAse activity from other sources, and by combinations thereof. Following such analyses, the priming oligonucleotide can be selected for so that it will hybridize to a section of RNA which is substantially nondegraded by the selective RNAse present in the reaction mixture, because substantial degradation at the binding site for the priming oligonucleotide could inhibit initiation of DNA synthesis and prevent optimal extension of the primer. In other words, a priming oligonucleotide is typically selected to hybridize to a region of an RNA target nucleic acid or the complement of a DNA target nucleic acid, located so that when the RNA is subjected to selective RNAse degradation, there is no substantial degradation which would prevent formation of the primer extension product.

Conversely, the site for hybridization of the promoter oligonucleotide may be chosen so that sufficient degradation of the RNA strand occurs to permit efficient hybridization of the promoter oligonucleotide to the DNA strand. Typically, only portions of RNA are removed from the RNA:DNA duplex through selective RNAse degradation and, thus, some parts of the RNA strand will remain in the duplex. Selective RNAse degradation on the RNA strand of an RNA:DNA hybrid results in the dissociation of small pieces of RNA from the hybrid. Positions at which RNA is selectively degraded may be determined through standard hybridization analyses. Thus, a promoter oligonucleotide may be selected which will more efficiently bind to the DNA after selective RNAse degradation, *i.e.*, will bind at areas where RNA fragments are selectively removed.

Figures 1A-1D and 2A-2D do not show the RNA portions which may remain after selective RNAse degradation. It is to be understood, however, that even though Figures 1A-1D and 2A-2D show complete removal of RNA from the DNA:RNA duplex, under certain conditions only partial removal actually occurs. Indeed, amplification as depicted in Figures 1A-1D and 2A-2D may be inhibited if a substantial portion of the RNA strand of an RNA:DNA hybrid remains undegraded, thus preventing hybridization of the promoter oligonucleotide and/or the optional extender oligonucleotide. However, based upon principles and methods disclosed in this application, as well as those disclosed by Kacian *et al*, U.S. Patent No. 5,339,491, routine modifications can be made by those skilled in the art according

to the teachings of this invention to provide an effective and efficient procedure for amplification of RNA.

In summary, the present invention provides methods for autocatalytically synthesizing multiple copies of a target sequence from a target nucleic acid without repetitive manipulation of reaction conditions such as temperature, ionic strength and pH, which comprises combining into a reaction mixture a target nucleic acid which comprises either an RNA target sequence, or a single-stranded or partially single-stranded DNA target sequence or a double-stranded DNA sequence which has been rendered at least partially single-stranded; a priming oligonucleotide, a promoter oligonucleotide, and, optionally, a displacer oligonucleotide, an extender oligonucleotide and/or a binding molecule or other means for terminating a primer extension reaction, all as described above; a reverse transcriptase or an RNA-dependent DNA polymerase and a DNA-dependent DNA polymerase; an enzyme activity which selectively degrades the RNA strand of an RNA:DNA complex (such as an RNAse H activity); and an RNA polymerase which recognizes the promoter sequence in the promoter oligonucleotide. The reaction mixture also includes the necessary building blocks for nucleic acid amplification, *e.g.*, nucleoside triphosphates, buffers, salts, and stabilizing agents. The components of the reaction mixture may be combined stepwise or at once. The reaction mixture is incubated under conditions whereby an oligonucleotide:target nucleic acid is formed, and DNA priming and nucleic acid synthesis can occur for a period of time sufficient to allow multiple copies of the target sequence or its complement to be produced. The reaction advantageously takes place under conditions suitable for maintaining the stability of reaction components, such as the enzymes, and without requiring modification or manipulation of reaction conditions during the course of the amplification reaction. Accordingly, the reaction of some embodiments may take place under conditions that are substantially isothermal and include substantially constant ionic strength and pH.

As such, the amplification methods of the present invention do not require repeated denaturation steps to separate the RNA:DNA complexes produced upon extension of the priming oligonucleotide. A denaturation step would require manipulation of reaction conditions, such as by substantially increasing the temperature of the reaction mixture (generally from ambient temperature to a temperature between about 80°C and about 105°C), reducing its ionic strength (generally by 10X or more) or changing pH (usually increasing pH to 10 or greater). Such manipulations of the reaction conditions often deleteriously affect enzyme activities, requiring addition of additional enzyme and also necessitate further manipulations of the reaction mixture to return it to conditions suitable for further nucleic acid synthesis. In those embodiments where the target nucleic acid is double-stranded DNA, an initial denaturation step is often required. Denaturation may be carried out by altering

temperature, ionic strength, and/or pH as described above, prior to adding the remaining components of the reaction mixture. Once the remaining components are added, no additional manipulations of the reaction mixture are needed.

The methods of the present invention are designed to decrease, diminish, or substantially eliminate side-product formation in the amplification reactions. For example, side-products are decreased, diminished, or substantially eliminated through the utilization of promoter oligonucleotides modified to prevent primer extension by a DNA polymerase, generally by including a blocking moiety at the 3'-termini of the promoter oligonucleotides. Further embodiments decrease, diminish, or substantially eliminate side-products through the use of a cap which hybridizes to a region at the 3'-end of the priming oligonucleotide, thereby preventing oligonucleotide dimer formation. According to the present invention, most, *e.g.*, at least about 90%, of the oligonucleotides present in the amplification reaction which comprise a promoter further comprise a 3'-blocking moiety to prevent primer extension. In a specific embodiment, any oligonucleotide used in the amplification reaction which comprises a promoter, not just the promoter oligonucleotide, further comprises a 3'-blocking moiety. In certain preferred embodiments, most, *e.g.*, at least about 80%, 90%, 95%, 96%, 97%, 98% or 99%, or all oligonucleotides required for the amplification reaction, other than the priming oligonucleotides (including displacer oligonucleotides, if used), comprise a 3'-blocking moiety. Thus, in certain embodiments, most if not all DNA polymerase activity in the amplification reactions is limited to the formation of DNA primer extension products which comprise the priming oligonucleotides.

Promoters or promoter sequences suitable for incorporation in promoter oligonucleotides used in the methods of the present invention are nucleic acid sequences (either naturally occurring, produced synthetically or a product of a restriction digest) that are specifically recognized by an RNA polymerase that recognizes and binds to that sequence and initiates the process of transcription, whereby RNA transcripts are produced. Typical, known and useful promoters include those which are recognized by certain bacteriophage polymerases, such as those from bacteriophage T3, T7, and SP6, and a promoter from *E. coli*. The sequence may optionally include nucleotide bases extending beyond the actual recognition site for the RNA polymerase which may impart added stability or susceptibility to degradation processes or increased transcription efficiency. Promoter sequences for which there is a known and available polymerase that is capable of recognizing the initiation sequence are particularly suitable to be employed.

Suitable DNA polymerases include reverse transcriptases. Particularly suitable DNA polymerases include AMV reverse transcriptase and MMLV reverse transcriptase. Some of the reverse transcriptases suitable for use in the methods of the present invention, such as

AMV and MMLV reverse transcriptases, have an RNAse H activity. Indeed, according to certain embodiments of the present invention, the only selective RNAse activity in the amplification reaction is provided by the reverse transcriptase -- no additional selective RNAse is added. However, in some situations it may also be useful to add an exogenous selective RNAse, such as *E. coli* RNAse H. Although the addition of an exogenous selective RNAse is not required, under certain conditions, the RNAse H activity present in, *e.g.*, AMV reverse transcriptase may be inhibited or inactivated by other components present in the reaction mixture. In such situations, addition of an exogenous selective RNAse may be desirable. For example, where relatively large amounts of heterologous DNA are present in the reaction mixture, the native RNAse H activity of the AMV reverse transcriptase may be somewhat inhibited and thus the number of copies of the target sequence produced accordingly reduced. In situations where the target nucleic acid comprises only a small portion of the nucleic acid present (*e.g.*, where the sample contains significant amounts of heterologous DNA and/or RNA), it is particularly useful to add an exogenous selective RNAse. *See, e.g.*, Kacian *et al*, U.S. Patent No. 5,399,491, the contents of which are hereby incorporated by reference herein (*see* Example 8).

RNA amplification products produced by the methods described above may serve as templates to produce additional amplification products related to the target sequence through the above-described mechanisms. The system is autocatalytic and amplification by the methods of the present invention occurs without the need for repeatedly modifying or changing reaction conditions such as temperature, pH, ionic strength and the like. These methods do not require an expensive thermal cycling apparatus, nor do they require several additions of enzymes or other reagents during the course of an amplification reaction.

The methods of the present invention are useful in assays for detecting and/or quantitating specific nucleic acid target sequences in clinical, environmental, forensic, and similar samples or to produce large numbers of RNA amplification products from a specific target sequence for a variety of uses. For example, the present invention is useful to screen clinical samples (*e.g.*, blood, urine, feces, saliva, semen, or spinal fluid), food, water, laboratory and/or industrial samples for the presence of specific nucleic acids. The present invention can be used to detect the presence of, for example, viruses, bacteria, fungi, or parasites. The present invention is also useful for the detection of human, animal, or plant nucleic acids for genetic screening, or in criminal investigations, archeological or sociological studies.

In a typical assay, a sample containing a target nucleic acid to be amplified is mixed with a buffer concentrate containing the buffer, salts, magnesium, triphosphates, oligonucleotides, *e.g.*, a priming oligonucleotide, a promoter oligonucleotide, and, optionally,

a displacer oligonucleotide, an extender oligonucleotide and/or a binding molecule, *e.g.*, a terminating oligonucleotide or a digestion oligonucleotide, and/or a capping oligonucleotide, and other reagents. The reaction may optionally be incubated at a temperature, *e.g.*, 60-100°C, for a period of time sufficient to denature any secondary structures in the target nucleic acid or to denature a double-stranded DNA target nucleic acid. After cooling, reverse transcriptase, an RNA polymerase, and, if desired, a separate selective RNAse, *e.g.*, RNAse H, are added and the reaction is incubated for a specified amount of time, *e.g.*, from about 10 minutes to about 120 minutes, at an optimal temperature, *e.g.*, from about 20 °C to about 55°C, or more, depending on the reagents and other reaction conditions. The displacer oligonucleotide, if used, may be provided with the enzymes to permit sufficient time for the priming oligonucleotide to hybridize to the target nucleic acid before initiating amplification.

The amplification product can be detected by hybridization with an optionally labeled detection probe and measurement of the resulting hybrids can be performed in any conventional manner. Design criteria in selecting probes for detecting particular target sequences are well known in the art and are described in, for example, Hogan *et al.*, "Methods for Making Oligonucleotide Probes for the Detection and/or Quantitation of Non-Viral Organisms," U.S. Patent No. 6,150,517, the contents of which are hereby incorporated by reference herein. Hogan teaches that probes should be designed to maximize homology for the target sequence(s) and minimize homology for possible non-target sequences. To minimize stability with non-target sequences, Hogan instructs that guanine and cytosine rich regions should be avoided, that the probe should span as many destabilizing mismatches as possible, and that the length of perfect complementarity to a non-target sequence should be minimized. Contrariwise, stability of the probe with the target sequence(s) should be maximized, adenine and thymine rich regions should be avoided, probe:target hybrids are preferably terminated with guanine and cytosine base pairs, extensive self-complementarity is generally to be avoided, and the melting temperature of probe:target hybrids should be about 2-10°C higher than the assay temperature.

In particular, the amplification product can be assayed by the Hybridization Protection Assay ("HPA"), which involves hybridizing a chemiluminescent oligonucleotide probe to the target sequence, *e.g.*, an acridinium ester-labeled ("AE") probe, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining probe in a luminometer. *See, e.g.*, Arnold *et al.*, "Homogenous Protection Assay," U.S. Patent No. 5,283,174 and NORMAN C. NELSON ET AL., NONISOTOPIC PROBING, BLOTTING, AND SEQUENCING, ch. 17 (Larry J. Kricka ed., 2d ed. 1995), each of which is hereby incorporated by reference in its entirety. Particular methods of carrying out HPA using AE probes are disclosed in the Examples section hereinbelow.

In further embodiments, the present invention provides quantitative evaluation of the amplification process in real-time by methods described herein. Evaluation of an amplification process in "real-time" generally involves making periodic determinations of the amount of signal associated with probe:amplicon complexes in the reaction mixture during the amplification reaction, and the determined values are used to calculate the amount of target sequence initially present in the sample. There are a variety of methods for determining the amount of initial target sequence present in a sample based on real-time amplification. These include those disclosed by Light *et al.*, "Method for Determining the Amount of an Analyte in a Sample," U.S. Patent Application Publication No. US 2006-0276972 (paragraphs 505 to 549); Lee *et al.*, "Methods for Quantitative Analysis of a Nucleic Acid Amplification Reaction," U.S. Patent Application Publication No. US 2006-0286587; Carrick *et al.*, "Method and Algorithm for Quantitating Polynucleotides," U.S. Patent Application Publication No. US 2006-0292619; Wittwer *et al.*, "Method for Quantification of an Analyte," U.S. Patent No. 6,303,305; and Yokoyama *et al.*, "Method for Assaying Nucleic Acid," U.S. Patent No. 6,541,205. (Each of these references or the indicated portion is hereby incorporated by reference herein.) Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-time amplification, is disclosed by Ryder *et al.*, "Method for Determining Pre-Amplification Levels of a Nucleic Acid Target Sequence from Post-Amplification Levels of Product," U.S. Patent No. 5,710,029, the contents of which are hereby incorporated by reference herein. The present invention is particularly suited to real-time evaluation, because the production of side-products is decreased, diminished, or substantially eliminated.

Amplification products may be detected in real-time through the use of various self-hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybridized state or an altered state through hybridization to a target sequence. By way of example, "molecular torches" are a type of self-hybridizing probe which includes distinct regions of self-complementarity (referred to as "the target binding domain" and "the target closing domain") which are connected by a joining region (*e.g.*, non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domains that are from 1 to about 10 bases in length and are accessible for hybridization to a target sequence present in an amplification product under strand displacement conditions. The single-stranded region may be, for example, a terminal region or an internal region, such as a loop region. Alternatively, the strand displacement conditions may cause "breathing" in a double-stranded terminal region of the molecular torch, thereby

resulting in a transient single-stranded region of the terminal region which is accessible for hybridization to the target sequence. Under strand displacement conditions, hybridization of the two complementary regions (which may be fully or partially complementary) of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular torch include a detectable label or a pair of interacting labels (*e.g.*, luminescent/quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe:target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and a variety of types of interacting label pairs are disclosed by Becker *et al.*, "Molecular Torches," U.S. Patent No. 6,534,274, the contents of which are hereby incorporated by reference herein.

Another example of a detection probe exhibiting self-complementarity is a "molecular beacon." Molecular beacons include nucleic acid molecules having a target complement sequence, an affinity pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification product, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complement sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (*e.g.*, DABCYL and EDANS). Molecular beacons are disclosed by Tyagi *et al.*, "Detectably Labeled Dual Confirmation Oligonucleotide Probes, Assays and Kits," U.S. Patent No. 5,925,517, and Tyagi *et al.*, "Nucleic Acid Detection Probes Having Non-FRET Fluorescence Quenching and Kits and Assays Including Such Probes," U.S. Patent No. 6,150,097, each of which is hereby incorporated by reference herein in its entirety.

Other self-hybridizing probes for use in the present invention are well known to those of ordinary skill in the art. By way of example, probe binding pairs having interacting labels, such as those disclosed by Morrison, "Competitive Homogenous Assay," U.S. Patent No. 5,928,862 (the contents of which are hereby incorporated by reference herein), might be adapted for use in the present invention. Probe systems used to detect single nucleotide polymorphisms (snps) might also be utilized in the present invention. Additional detection systems include "molecular switches," as disclosed by Arnold *et al.*, "Oligonucleotides Comprising a Molecular Switch," U.S. Patent Application Publication No. US 2005-0042638, the contents of which are hereby incorporated by reference herein. And other probes, such as those comprising intercalating dyes and/or fluorochromes, might be useful for detection of

amplification products in the present invention. See, *e.g.*, Ishiguro *et al.*, "Method of Detecting Specific Nucleic Acid Sequences," U.S. Patent No. 5,814,447, the contents of which are hereby incorporated by reference herein.

In those methods of the present invention where the initial target sequence and the RNA transcription product share the same sense, it may be desirable to initiate amplification before adding probe for real-time detection. Adding probe prior to initiating an amplification reaction may slow the rate of amplification since probe which binds to the initial target sequence has to be displaced or otherwise remove during the primer extension step to complete a primer extension product having the complement of the target sequence. The initiation of amplification is judged by the addition of amplification enzymes (*e.g.*, a reverse transcriptase and an RNA polymerase).

In addition to the methods described herein, the present invention is drawn to kits comprising one or more of the reagents required for carrying out the methods of the present invention. Kits comprising various components used in carrying out the present invention may be configured for use in any procedure requiring amplification of nucleic acid target molecules, and such kits can be customized for various different end-users. Suitable kits may be prepared, for example, for blood screening, disease diagnosis, infection control, environmental analysis, criminal investigations or other forensic analyses, genetic analyses, archeological or sociological analyses, or for general laboratory use. Kits of the present invention provide one or more of the components necessary to carry out nucleic acid amplifications according to the invention. Kits may include reagents suitable for amplifying nucleic acids from one particular target or may include reagents suitable for amplifying multiple targets. Kits of the present invention may further provide reagents for real-time detection of one or more nucleic acid targets in a single sample, for example, one or more self-hybridizing probes as described above. Kits may comprise a carrier that may be compartmentalized to receive in close confinement one or more containers such as vials, test tubes, wells, and the like. Preferably at least one of such containers contains one or more components or a mixture of components needed to perform the amplification methods of the present invention.

A kit according to the present invention can include, for example, in one or more containers, a priming oligonucleotide, a promoter oligonucleotide modified to prevent primer extension by a DNA polymerase (*e.g.*, modified to include a 3'-blocking moiety), a binding molecule or other means for terminating a primer extension reaction, and, optionally, a displacer oligonucleotide, an extender oligonucleotide and/or a capping oligonucleotide as described herein. If real-time detection is used, the one or more containers may include one or more reagents for real-time detection of at least one nucleic acid target sequence in a single

sample, for example, one or more self-hybridizing probes as described above. Another container may contain an enzyme reagent, for example a mixture of a reverse transcriptase (either with or without RNAse H activity), an RNA polymerase, and optionally an additional selective RNAse enzyme. (If included, the displacer oligonucleotide may be provided in the container containing the enzyme reagent.) These enzymes may be provided in concentrated form or at working concentration, usually in a form which promotes enzyme stability. The enzyme reagent may also be provided in a lyophilized form. *See* Shen *et al.*, "Stabilized Enzyme Compositions for Nucleic Acid Amplification," U.S. Patent No. 5,834,254, the contents of which are hereby incorporated by reference herein. Another one or more containers may contain an amplification reagent in concentrated form, *e.g.*, 10X, 50X, or 100X, or at working concentration. An amplification reagent will contain one or more of the components necessary to run the amplification reaction, *e.g.*, a buffer, $MgCl_2$, KCl, dNTPs, rNTPs, EDTA, stabilizing agents, *etc.* Certain of the components, *e.g.*, $MgCl_2$ and rNTPs, may be provided separately from the remaining components, allowing the end user to titrate these reagents to achieve more optimized amplification reactions. Another one or more containers may include reagents for detection of amplification products, including one or more optionally labeled detection probes. In some embodiments, a kit of the present invention will also include one or more containers containing one or more positive and negative control target nucleic acids which can be utilized in amplification experiments in order to validate the test amplifications carried out by the end user. In some instances, one or more of the reagents listed above may be combined with an internal control. It is also possible to combine one or more of these reagents in a single tube or other containers.

Supports suitable for use with the invention (*e.g.*, test tubes, multi-tube units, multi-well plates, cuvettes, flexible containers, microfluidic devices, including analytical cards or discs for use in centrifugal analyzers, etc.) may also be supplied with reagents of the invention. Finally, a kit of the present invention may include one or more instruction manuals provided in written or electronic form, including CD-ROMs, DVDs and video tapes. Kits of the invention may contain virtually any combination of the components set out above or described elsewhere herein. As one skilled in the art would recognize, the components supplied with kits of the invention will vary with the intended use for the kits, and the intended end user. Thus, kits may be specifically designed to perform various functions set out in this application and the components of such kits will vary accordingly.

The present invention is further drawn to various oligonucleotides, including the priming oligonucleotides, promoter oligonucleotides, terminating oligonucleotides, displacer oligonucleotides, extender oligonucleotides, capping oligonucleotides and detection probes described herein. It is to be understood that the oligonucleotides of the present invention may

be DNA, RNA, DNA:RNA chimerics and analogs thereof, and, in any case, the present invention includes RNA equivalents of DNA oligonucleotides and DNA equivalents of RNA oligonucleotides. Except for the preferred priming oligonucleotides (including displacer oligonucleotides) and detection probes described *supra*, the oligonucleotides described hereinabove preferably comprise a blocking moiety at their 3'-termini.

Detection probes of the present invention may be labeled in a number of alternative ways, *e.g.*, with radioactive isotopes, fluorescent labels, chemiluminescent labels, nuclear tags, bioluminescent labels, intercalating dyes, or enzyme labels. The detection probes may include groups of interacting labels which emit a detectable change in signal when the probes are hybridized to a target sequence or its complement. In various embodiments, these labeled probes optionally or preferably are synthesized to include at least one modified nucleotide, *e.g.*, a 2'-O-methyl ribonucleotide; or these labeled oligonucleotide probes optionally or preferably are synthesized entirely of modified nucleotides, *e.g.*, 2'-O-methyl ribonucleotides.

It will be understood by one of ordinary skill in the relevant arts that other suitable modifications and adaptations to the methods and compositions described herein are readily apparent from the description of the invention contained herein in view of information known to the ordinarily skilled artisan, and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

EXAMPLES

Examples are provided below illustrating different aspects and embodiments of the invention. It is believed that these examples accurately reflect the details of experiments actually performed, however, it is possible that some minor discrepancies may exist between the work actually performed and the experimental details set forth below which do not affect the conclusions of these experiments. Skilled artisans will appreciate that these examples are not intended to limit the invention to the specific embodiments described therein. Additionally, those skilled in the art, using the techniques, materials and methods described herein, could easily devise and optimize alternative amplification systems for detecting and/or quantifying any target sequence.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, recombinant DNA, and chemistry, which are within the skill of the art. Such techniques are explained fully in the literature. *See*, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., Sambrook *et al.*, ed., Cold

Spring Harbor Laboratory Press: (1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis *et al.* U.S. Pat. No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); and in Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1989).

Unless otherwise indicated, oligonucleotides and modified oligonucleotides in the following examples were synthesized using standard phosphoramidite chemistry, various methods of which are well known in the art. See *e.g.*, Carruthers, et al., 154 Methods in Enzymology, 287 (1987), the contents of which are hereby incorporated by reference herein. Unless otherwise stated herein, modified nucleotides were 2'-O-methyl ribonucleotides, which were used in the synthesis as their phosphoramidite analogs. Applicant prepared the oligonucleotides using an Expedite™ 8909 DNA Synthesizer (PerSeptive Biosystems, Framingham, Mass.).

Various reagents are identified in the examples below, which include an amplification reagent, an enzyme reagent, a hybridization reagent, a selection reagent, and detection reagents. Unless otherwise indicated, the formulations and pH values (where relevant) of these reagents were as follows.

Amplification Reagent. The "Amplification Reagent" comprised 11.6 mM Trizma® base buffer, 15 mM Trizma® hydrochloride buffer, 22.7 mM $MgCl_2$, 23.3 mM KCl, 3.33% (v/v) glycerol, 0.05 mM zinc acetate, 0.665 mM dATP, 0.665 mM dCTP, 0.665 mM dGTP, 0.665 mM dTTP, 0.02% (v/v) ProClin 300 Preservative (Supelco, Bellefonte, PA; Cat. No. 48126), 5.32 mM ATP, 5.32 mM CTP, 5.32 mM GTP, 5.32 mM UTP, and 6 M HCl to pH 7.81 to 8.0 at 22 °C.

Enzyme Reagent. The "Enzyme Reagent" comprised 70 mM N-acetyl-L-cysteine, 10% (v/v) TRITON® X-102 detergent, 16 mM HEPES, 3 mM EDTA, 0.05% (w/v) sodium azide, 20 mM Trizma® base buffer, 50 mM KCl, 20% (v/v) glycerol, 150 mM trehalose, 4M NaOH to pH 7, and containing 224 RTU/μL Moloney murine leukemia virus ("MMLV") reverse transcriptase and 140 U/μL T7 RNA polymerase, where one RTU of RT activity incorporates 1 nmol of dTMP into DE81 filter-bound product in 20 minutes at 37°C using poly(rA)-$p(dT)_{12-18}$ as the substrate, and one U of T7 RNA polymerase activity produces 5 fmol of RNA transcript in 20 minutes at 37°C.

Hybridization Reagent. The "Hybridization Reagent" comprised 100 mM succinic acid, 2% (w/v) lithium lauryl sulfate, 230 mM LiOH, 15 mM aldrithiol-2, 1.2 M LiCl, 20 mM EDTA, 20 mM EGTA, 3.0% (v/v) ethyl alcohol, and 2M LiOH to pH 4.7.

Selection Reagent. The "Selection Reagent" comprised 600 mM H_3BO_3, 182 mM NaOH, 1% (v/v) TRITON® X-100 detergent, and 4 M NaOH to pH 8.5.

Detection Reagent I. "Detection Reagent I" comprised 1 mM HNO_3 and 30 mM H_2O_2.

Detection Reagent II. "Detection Reagent II" comprised 1 M NaOH and 2% (w/v) Zwittergent® 3-14 detergent.

Oil Reagent: The "Oil Reagent" comprised a silicone oil (United Chemical Technologies, Inc., Bristol, PA; Cat. No. PS038).

Example 1
Comparison of Blocked and Unblocked Promoter Oligonucleotides

This experiment was conducted to evaluate the specificity of an amplification method according to the present invention in which a region ("the target region") of a cloned transcript derived from the 5' untranslated region of the hepatitis C virus ("the transcript") was targeted for amplification. For this experiment we prepared two sets of priming and promoter oligonucleotides having identical base sequences. The two sets of oligonucleotides differed by the presence or absence of a 3'-terminal blocking moiety on the promoter oligonucleotide. The promoter oligonucleotide in each set targeted the complement of a sequence contained within the 5'-end of the target region and had the base sequence of SEQ ID NO:5 aatttaatacgactcactatagggagactagccatg gcgttagtatgagtgtcgtgcag, where the underlined portion of the promoter oligonucleotide constitutes a T7 promoter sequence (SEQ ID NO:3) and the non-underlined portion represents a hybridizing sequence (SEQ ID NO:4). The priming oligonucleotide in each set targeted a sequence contained within the 3'-end of the target region and had the base sequence of SEQ ID NO:6. Also included in the amplification method was a terminating oligonucleotide made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:38 ggcuagacgcuuucugcgugaaga. The terminating oligonucleotide had a 3'-terminal blocking moiety and targeted a region of the transcript just 5' to the target region. The 5'-ends of the terminating oligonucleotide and of the hybridizing sequence of the promoter oligonucleotide overlapped by six bases. The 3'-terminal blocking moiety of both

the promoter oligonucleotide and the terminating oligonucleotide consisted of a 3'-to-3' linkage prepared using 3'-dimethyltrityl-N-benzoyl-2'-deoxycytidine, 5'-succinoyl-long chain alkylamino-CPG (Glen Research Corporation, Sterling, VA; Cat. No. 20-0102-01).

For amplification, 75 μL of the Amplification Reagent was added to each of eight reaction tubes. The Amplification Reagent was then combined with 30 pmol of a promoter oligonucleotide, 30 pmol of the priming oligonucleotide and 5 pmol of the terminating oligonucleotide. One set of four of the tubes was provided with 30 pmol of the unblocked promoter oligonucleotide (group I), and another set of four tubes was provided with 30 pmol of the blocked promoter oligonucleotide (group II). Next, 1 μL of a 0.1% (w/v) lithium lauryl sulfate ("LLS") buffer containing 1000 copies/μL of the transcript was added to two of the tubes in each group, while the remaining two tubes in each group served as negative controls. The reaction mixtures were overlaid with 200 μL of the Oil Reagent, and the tubes were then sealed and hand-shaken horizontally for 5 to 10 seconds before the tubes were incubated in a 60°C water bath for 10 minutes. The tubes were then transferred to a 41.5°C water bath and incubated for 15 minutes before adding 25 μL of the Enzyme Reagent to each tube. After adding the Enzyme Reagent, the tubes were sealed, removed from the water bath and hand-shaken horizontally for 5 to 10 seconds to fully mix the components of the reaction mixtures. The tubes were returned to the 41.5°C water bath and incubated for an additional 60 minutes to facilitate amplification of the target region in the presence of MMLV reverse transcriptase and T7 RNA polymerase. Following amplification, the tubes were removed from the 41.5°C water bath and allowed to cool at room temperature for 10 to 15 minutes.

A 5 μL aliquot of each reaction mixture was taken from the tubes, diluted 1:1 with a 2X Novex® TBE-Urea Sample Buffer (Invitrogen Corporation, Carlsbad, CA; Cat. No. LC6876), and loaded onto a Novex® TBE-Urea Denaturing Gel (Invitrogen; Cat. No. EC6865BOX). The gel was held by an Xcell Surelock™ Mini-Cell (Invitrogen; Cat. No. EI0001) and run at 180 volts for 50 minutes using a 5X Novex® TBE Running Buffer (Invitrogen; Cat. No. LC6675) diluted 1:4 with deionized water. Afterwards, the gel was stained with 0.5 μg/mL of ethidium bromide in a 1X TBE (Tris-Borate-EDTA) solution, visualized on a FisherBiotech® Ultraviolet Transilluminator (Fisher Scientific International Inc., Hampton, NH; Model No. FB-TIV-816A), and photographed with a handheld camera using Polaroid 667 film.

The results of this experiment are illustrated in the photographed gel of FIG. 3. Each number above the pictured gel represents a distinct lane, where lane 1 is an RNA ladder of 100, 200, 300, 400, 500, 750 and 1000 base oligonucleotides, and the remainder of the lanes correspond to the following reaction mixtures: (i) lanes 2 and 3 correspond to the transcript-containing replicates of group I (unblocked promoter oligonucleotide); (ii) lanes 4 and 5

correspond to the transcript-containing replicates of group II (blocked promoter oligonucleotide); (iii) lanes 6 and 7 correspond to the transcript-negative replicates of group I; and (iv) lanes 8 and 9 correspond to the transcript-negative replicates of group II. The first visible band in lanes 2-5 constitutes amplicon derived from amplification of the target region. The remainder of the bands in lanes 2, 3, 6 and 7 constitute non-specific amplification products. Thus, the results indicate that only amplification using the fully blocked promoter oligonucleotides was specific, as there was no visible side-product formation in either the transcript-containing or transcript negative reaction mixtures containing blocked promoter oligonucleotides, whereas visible side-products were formed in both the transcript-containing and transcript-negative reaction mixtures containing unblocked promoter oligonucleotides.

Example 2
Reduction in the Formation of Replicating Molecules

This experiment was designed to evaluate whether the use of a blocked promoter oligonucleotide in an amplification method of the present invention would lead to a reduction in the formation of replicating molecules over a standard transcription-based amplification procedure. Replicating molecules are generally believed to form when the 3'-end of a promoter oligonucleotide forms a hairpin structure and is extended in the presence of a polymerase, thereby forming a double-stranded promoter sequence. Transcription initiated from the double-stranded promoter sequence results in the formation of amplicon containing an antisense version of the promoter sequence.

In this experiment, we compared the production of replicating molecules in amplification reactions containing promoter oligonucleotides that were either blocked or unblocked at their 3'-terminal ends in the presence or absence of purified rRNA from *Mycobacterium tuberculosis* (ATCC No. 25177) using one of two detection probes targeting a region ("the target region") of the 16S rRNA of *Mycobacterium tuberculosis* ("the target nucleic acid"). The blocked and unblocked promoter oligonucleotides targeted sequences contained within the complement of the 5'-end of the target region. The blocked promoter oligonucleotide had the base sequence of SEQ ID NO:26 aattctaatacgactcactatagggagaactgggtctaataccggataggaccacgggatgcat, and the unblocked promoter oligonucleotide had the base sequence of SEQ ID NO:28 aattctaatacgactcactatagggagaactgggtctaat accggataggaccacgggga, where the underlined portion of each promoter oligonucleotide constitutes a T7 promoter sequence (SEQ ID NO:3) and the non-underlined portion represents a hybridizing sequence (SEQ ID NO:25 and SEQ ID NO:27). The priming oligonucleotide targeted a sequence contained within the 3'-end of the target region and had the base sequence of SEQ ID NO:29. Also included was a terminating

oligonucleotide made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:39 cccaguuucccaggcuuauccc. The terminating oligonucleotide targeted a region of the target nucleic acid just 5' to the target region and had a 3'-terminal blocking moiety. The 5'-ends of the terminating oligonucleotide and the hybridizing sequence of the promoter oligonucleotide overlapped by six bases. The 3'-terminal blocking moiety of both the blocked promoter oligonucleotide and the terminating oligonucleotide consisted of the 3'- to-3' linkage described in Example 1. And for detection, two detection probes were synthesized. The first detection probe ("detection probe I") comprised 2'-O-methyl ribonucleotides targeted a sequence contained within the target region and had the base sequence of SEQ ID NO:30 gcucauccca*caccgcuaaagc. The second detection probe ("detection probe II") targeted the antisense of a region contained within the T7 promoter sequence and had the base sequence of SEQ ID NO:40 atacgactc*actata. The asterisk in both detection probe sequences indicates the position of a 4-(2-succinimidyloxycarbonyl ethyl)-phenyl-10-methylacridinium-9-carboxylate fluorosulfonate acridinium ester label ("standard AE") joined to the probe by means of a non-nucleotide linker, as described by Arnold *et al.*, "Linking Reagents for Nucleotide Probes," U.S. Patent No. 5,585,481, the contents of which are hereby incorporated by reference herein.

A total of eight different amplification reactions were performed in replicates of five. All of the reaction tubes used for the amplification reactions were provided with 75 µL of the Amplification Reagent, followed by 5 pmol each of either the blocked or unblocked promoter oligonucleotide, the priming oligonucleotide, and the terminating oligonucleotide. Two sets of the tubes received 2 µL each of a 0.2% (w/v) LLS buffer containing 250 copies/µL of the target nucleic acid, and the other two sets of tubes received no target nucleic acid. The reaction mixtures were overlaid with 200 µL of the Oil Reagent, and the tubes were then sealed and hand-shaken horizontally for 5 to 10 seconds before being incubated in a 60°C water bath for 10 minutes. The tubes were then transferred to a 41.5°C water bath and incubated for 10 minutes before adding 25 µL of the Enzyme Reagent to each tube. After adding the Enzyme Reagent, the tubes were sealed, removed from the water bath and hand-shaken horizontally for 5 to 10 seconds to fully mix the components of the reaction mixtures. The tubes were returned to the 41.5°C water bath and incubated for an additional 60 minutes to permit amplification of the target sequences. Following amplification, the tubes were removed from the 41.5°C water bath and allowed to cool at room temperature for 10 to 15 minutes.

The detection step was performed in accordance with the Hybridization Protection Assay disclosed by Arnold *et al.*, "Homogenous Protection Assay," U.S. Patent No. 5,283,174. In this step, 100 µL of the Hybridization Reagent containing either 52 fmol of

detection probe I or 10.2 fmol of detection probe II was added to each tube. After adding the detection probes, the tubes were sealed, hand-shaken horizontally for 5 to 10 seconds, and incubated in a 60°C water bath for 15 minutes to permit hybridization of the detection probes to their corresponding target sequences. Following hybridization, 250 µL of the Selection Reagent was added to the tubes and the tubes were sealed and hand-shaken horizontally for 5 to 10 seconds before being incubated in a 60°C water bath for 10 minutes to hydrolyze acridinium ester labels associated with unhybridized probe. The samples were cooled at room temperature for at least 10 minutes before being analyzed in a LEADER® HC+ Luminometer (Gen-Probe Incorporated, San Diego, CA; Cat. No. 4747) equipped with automatic injection of Detection Reagent I, followed by automatic injection of Detection Reagent II. Signal emitted from the tubes was measured in relative light units ("RLU"), which is a measure of chemiluminescence.

The results were averaged for each set of reaction conditions and are presented in Table 1 below. From these results, it can be seen that those amplification reactions containing the blocked promoter oligonucleotide performed as well as those amplification reactions containing the unblocked promoter oligonucleotide at amplifying a targeted region of the target nucleic acid. However, those amplification reactions containing the blocked promoter oligonucleotide produced substantially fewer replicating molecules than did those amplification reactions containing the unblocked promoter oligonucleotide, both in the presence and in the absence of the transcript.

Table 1

Effect of 3'-Blocking Promoter Oligonucleotides on the Formation of Replicating Molecules

	Detection Probe I		Detection Probe II	
	Target Nucleic Acid	No Target Nucleic Acid	Target Nucleic Acid	No Target Nucleic Acid
Blocked Promoter Oligonucleotide	1,047,084	4,222	64,874	10,063
Unblocked Promoter Oligonucleotide	976,156	98,067	526,657	456,130

Example 3

Sensitivity of Amplification Assay Using Blocked Promoter Oligonucleotide and Terminating Oligonucleotide

This experiment examined the sensitivity of an amplification system according to the present invention in which a region ("the target region") of purified 23S rRNA from *Chlamydia trachomatis* (ATCC No. VR-878) ("the target nucleic acid") was targeted for amplification. Included in this experiment was a promoter oligonucleotide having a 3'-terminal blocking moiety, a priming oligonucleotide, a terminating oligonucleotide having a 3'-terminal blocking moiety, and a labeled detection probe. The promoter oligonucleotide targeted the complement of a sequence contained within the 5'-end of the target region and had the base sequence of SEQ ID NO:22 aatttaatacgactcactatagggagacggagtaagttaagcacgcggac gattgga, where the underlined portion of the promoter oligonucleotide constitutes a T7 promoter sequence (SEQ ID NO:3) and the non-underlined portion represents a hybridizing sequence (SEQ ID NO:21). The priming oligonucleotide targeted a sequence contained within the 3'-end of the target region and had the base sequence of SEQ ID NO:23. The terminating oligonucleotide was made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:41 uccgucauuccuucgcuauagu and targeted a region of the target nucleic acid just 5' to the target region. The 5'-ends of the terminating oligonucleotide and the hybridizing sequence of the promoter oligonucleotide overlapped by four bases. The 3'-terminal blocking moiety of both the promoter oligonucleotide and the terminating oligonucleotide consisted of the 3'-to-3' linkage described in Example 1. The detection probe targeted a sequence contained within the target region and was made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:24 cguucucaucgcucu*acggacucu, where the asterisk indicates the position of a standard AE label joined to the probe by means of a non-nucleotide linker. *See* Arnold *et al.*, U.S. Patent No. 5,585,481.

Amplification in this experiment was carried out essentially as described in Example 1. Each amplification reaction was performed in replicates of 3, and the target nucleic acid was added to each reaction tube in each set of replicates in copy numbers of 10, 100, 1000 or 10,000 from a 0.1% (w/v) LLS buffer containing 10, 100, 1000 or 10,000 copies/μL, respectively. The promoter and priming oligonucleotides were each added to the tubes in 30 pmol/reaction amounts, and 5 pmol of the terminating oligonucleotide was added to each tube. Using the *Chlamydia trachomatis* probe of this experiment, detection was carried out essentially as described in Example 2. The results of this experiment are set forth in Table 2 below and indicate 100 copy sensitivity for this amplification system, where an average RLU value of above 10,000 constituted a positive result.

Table 2
Sensitivity of *Chlamydia trachomatis* Amplification System

Copy Number	Avg. RLU
10	8504
100	51,574
1000	1,578,416
10,000	6,092,697

Example 4
Amplification of a Double-Stranded Target Sequence

This example examines an amplification system according to the present invention in which a region ("the target region") of a cloned, double-stranded transcript derived from the E6 and E7 genes of human papilloma virus type 16 ("HPV-16") ("the transcript") was targeted for amplification. *See* FIG. 1C. This experiment included a promoter oligonucleotide having a 3'-terminal blocking moiety, a priming oligonucleotide and a labeled detection probe. The promoter oligonucleotide targeted the complement of a sequence contained within the 5'-end of the target region and had the base sequence of SEQ ID NO:14 aatttaatacgactcactatagggagagaacagatggggcacacaattcctagt, where the underlined portion of the promoter oligonucleotide constitutes a T7 promoter sequence (SEQ ID NO:3) and the non-underlined portion represents a hybridizing sequence (SEQ ID NO:13). The 3'-terminal blocking moiety of the promoter oligonucleotide consisted of the 3'-to-3' linkage described in Example 1. The priming oligonucleotide targeted a sequence contained within the 3'-end of the target region and had the base sequence of SEQ ID NO:15. The detection probe, which was comprised of 2'-O-methyl ribonucleotides, had the base sequence of SEQ ID NO:16 ggacaa*gcagaaccggaca and targeted a sequence contained within the target region. The asterisk indicates the position of a standard AE label joined to the probe by means of a non-nucleotide linker. *See* Arnold *et al.*, U.S. Patent No. 5,585,481.

The amplification reactions of this experiment were performed in replicates of 5, and each tube included 75 µL of the Amplification Reagent containing 0, 50, 100, 500, 1000 or 5000 copies of the transcript. Each tube was also provided with 40 pmol of the promoter oligonucleotide and 15 pmol of the priming oligonucleotide. The reaction mixtures were overlaid with 200 µL of the Oil Reagent, and the tubes were then sealed and hand-shaken horizontally for 5 to 10 seconds. To separate the complementary strands of the double-

stranded transcript, the tubes were incubated in a heat block for 10 minutes at 95°C. At the end of this incubation, the tubes were removed from the heat block and rapidly cooled on ice for 5 minutes to promote association of the priming oligonucleotide and the targeted region of the transcript. The tubes were then incubated in a 41.5°C water bath for 10 minutes. To initiate amplification, 25 µL of the Enzyme Reagent was added to the tubes, which were then sealed and hand-shaken horizontally for 5 to 10 seconds to fully mix the Amplification and Enzyme Reagents. Amplification was then carried out by returning the tubes to the 41.5°C water bath for a 1 hour incubation.

Following amplification, detection of the amplification products was performed in the manner described in Example 2 using 100 fmol/reaction of the detection probe. The results of this experiment are set forth in Table 3 below and indicate 500 copy sensitivity for this amplification system, where an RLU value of 10,000 or greater constituted a positive result.

Table 3
Sensitivity of HPV-16 Amplification System

Copy Number	Avg. RLU	% Positive Amplifications
0	5410	0
50	5647	0
100	6018	0
500	19,928	80
1000	200,072	80
5000	371,641	100

Example 5
Comparison of Blocked and Unblocked Promoter Oligonucleotides

The purpose of this experiment was to evaluate the benefit of including a terminating oligonucleotide in the HCV amplification system of Example 1. *See* FIG. 1A. For this experiment, four different reaction mixtures were set up in replicates of 10 containing either 0 or 10 copies of the transcript of Example 1 in the presence or absence of a terminating oligonucleotide. The promoter, priming and terminating oligonucleotides were identical to those used in Example 1. Unlike Example 1, this experiment included two detection probes, both of which were made up of 2'-O-methyl ribonucleotides and targeted a sequence contained within the region of the transcript targeted for amplification. The first detection

probe had the base sequence of SEQ ID NO:7 guacu*caccgguucc, and the second detection probe had the base sequence of SEQ ID NO:8 agaccacua*uggcucucccggg. Each detection probe had a "cold," or unlabeled version, and a "hot," or labeled version. (Cold probes were used in this experiment to prevent saturation of the hot probes in the presence of a vast excess of amplicon, thereby permitting the extent of amplification to be evaluated.) The asterisks indicate the positions of standard AE labels joined to the hot probes by means of non-nucleotide linkers. *See* Arnold *et al.*, U.S. Patent No. 5,585,481.

The amplification reactions were essentially carried out in the manner described in Example 2 using 30 pmol/reaction of the promoter oligonucleotide and 15 pmol/reaction of each of the priming and terminating oligonucleotides. Detection was performed as described in Example 2 using 100 fmol/reaction of each of the two hot probes and each of the two cold probes in amounts corresponding to the ratios indicated in Table 4 below. The averaged results are set forth in Table 4 in relative light units ("RLU") and demonstrate that only those reaction mixtures containing the terminating oligonucleotide had 10 copy level sensitivity in the HCV amplification system. The coefficient of variation values ("%CV") appearing in Table 4 for the different copy levels tested constitute the standard deviation of the replicates over the mean of the replicates as a percentage.

Table 4

Sensitivity of the HCV Amplification System in the Presence and Absence of a Terminating Oligonucleotide

Copy Number	Terminating Oligonucleotide	Cold Prot/Hot Probe Ratio	Avg. RLU	%CV
0	+	25:1	15,813	7.5
10	+	25:1	635,695	83.5
0	-	5:1	15,378	14.3
10	-	5:1	20,730	37.5

Example 6

Varying Length of Base Overlap Between Promoter Oligonucleotide and Terminating Oligonucleotide

In this experiment, we studied the effect of varying the length of overlap between a blocked promoter oligonucleotide and a terminating oligonucleotide on amplification efficiency in the HCV amplification system of Example 1. The reaction mixtures were set up in replicates of four and each set was provided with 0 or 50 copies of the transcript of

Example 1. The amount of overlap between the promoter oligonucleotide and the terminating oligonucleotide, if present, was 2, 4 or 6 bases for each set of reaction mixtures. The promoter oligonucleotide, the priming oligonucleotide, and the detection probes were identical to those used in Example 5. The cold probes and hot probes were used at a ratio of 4:1. The three terminating oligonucleotides of this experiment were made up of 2'-O-methyl ribonucleotides and had the following base sequences: (i) SEQ ID NO:42 agacgcuuucugcgugaagacagu (2 base overlap); (ii) SEQ ID NO:43 cuagacgcuuucugcgugaagaca (4 base overlap); and (iii) SEQ ID NO:38 (6 base overlap).

The amplification reactions were carried out in reaction tubes in the manner described in Example 5 using 30 pmol/reaction of the promoter oligonucleotide and 15 pmol/reaction each of the priming oligonucleotide and the terminating oligonucleotides. Detection was performed as described in Example 2 using 100 fmol/reaction of each of the two hot probes and 400 fmol/reaction of each of the two cold probes. The averaged results are set forth in Table 5 in relative light units ("RLU") and indicate that under the conditions tested, six bases of overlap between the promoter oligonucleotide and the terminating oligonucleotide is optimal for the HCV amplification system. The skilled artisan could apply this method to any amplification system to determine the optimal amount of overlap between a promoter oligonucleotide and a terminating oligonucleotide using nothing more than routine experimentation.

Table 5

Effect of Terminating Oligonucleotide/Promoter Oligonucleotide Base Overlap on Amplification Efficiency

Copy Number	Terminating Oligonucleotide	Base Overlap	Avg. RLU
0	-	N/A	29,593
	+	2	25,430
	+	4	27,128
	+	6	27,732
50	-	N/A	265,250
	+	2	339,833
	+	4	253,577
	+	6	1,904,911

Example 7

Comparison of Real-Time Amplification Assays in the Presence or Absence of a Terminating Oligonucleotide

This experiment was conducted to determine whether a terminating oligonucleotide improves amplification performance in a real-time amplification assay. For this experiment, we used the *Mycobacterium tuberculosis* amplification system of Example 2, which included the unblocked promoter oligonucleotide having the base sequence of SEQ ID NO:28, the priming oligonucleotide having the base sequence of SEQ ID NO:29, and the blocked terminating oligonucleotide having the base sequence of SEQ ID NO:39. Also included was a molecular beacon detection probe having the base sequence of SEQ ID NO:31. The detection probe was synthesized to include a BHQ-2 Black Hole Quencher™ Dye joined to its 3'-end using a BHQ-2 Glycolate CPG (Biosearch Technologies, Inc., Novato, CA; Cat. No. CG5-5042G-1) and a Cy™5 Dye joined to its 5'-end using a Cy™5-CE phosphoramidite (Glen Research; Cat. No. 105915-90). The reactions were run in the wells of a Thermo Labsystems White Cliniplate 96 (VWR International, Inc., West Chester, PA; Cat. No. 28298-610), and each reaction well contained 0, 100 or 1000 copies of the target nucleic acid of Example 2. For each copy number tested, there were four replicates which included the terminating oligonucleotide and four replicates which did not.

For amplification and detection, 75 µL of the Amplification Reagent was added to each reaction well, followed by the addition of 2 µL of a 0.1% (w/v) LLS buffer containing 50 copies/µL to each tube of one set of replicates and 2 µL of a 0.1% (w/v) LLS buffer containing 500 copies/µL to each tube of another set of replicates. The promoter oligonucleotide, the priming oligonucleotide and, when included, the terminating oligonucleotide were each added to the tubes in 5 pmol/reaction amounts, and 2 pmol/reaction of the detection probe was added to each tube. Target nucleic acid was provided to the reaction wells in the amounts indicated, and the reactions mixtures were overlaid with 80 µL of the Oil Reagent. The plate was sealed with a ThermalSeal RT™ Film (Sigma-Aldrich Corporation, St. Louis, MO; Product No. Z369675) and the contents of the plate were subjected to a 60°C incubation for 15 minutes in a Solo HT Microplate Incubator (Thermo Electron Corporation, Waltham, MA; Model No. 5161580), followed by a 42°C incubation for 10 minutes in the Solo HT Microplate Incubator. Next, 25 µL of the Enzyme Reagent (pre-heated to 42°C) was added to each well and the contents were mixed several times using a pipette. The contents of the plate were then incubated at 42°C for 120 minutes in a Biolumin™ 960 Micro Assay Reader (Molecular Dynamics Inc., Sunnyvale, CA) and fluorescence from the Cy™5 Dye channel was monitored as a function of time in one minute intervals. The results of this monitoring, which are graphically presented in Figures 4A-F,

indicate that the terminating oligonucleotide dramatically enhanced amplification of the target sequence in the *Mycobacterium tuberculosis* real-time amplification assay.

Example 8

Terminating Oligonucleotides Versus Digestion Oligonucleotides

This experiment compared levels of amplification in the *Mycobacterium tuberculosis* amplification system of Example 2 using either a terminating oligonucleotide or a digestion oligonucleotide in the presence of a blocked or unblocked promoter oligonucleotide. The terminating oligonucleotide of this experiment was designed to bind to the targeted RNA and physically block the activity of the reverse transcriptase enzyme, while the digestion oligonucleotide, which was composed of DNA, was designed to bind to the targeted RNA and direct digestion of the substrate RNA by an RNAse H activity. Use of the terminating or digestion oligonucleotide results in the formation of a template-complementary strand, or cDNA, having a defined 3'-end. The promoter oligonucleotide is designed so that its template-binding portion hybridizes to a 3'-terminal sequence present in the template-complementary strand, thereby facilitating the formation of a double-stranded promoter sequence in the presence of the reverse transcriptase enzyme.

As in Example 2, the promoter oligonucleotide of this experiment had the base sequence of SEQ ID NO:28 and the priming oligonucleotide had the base sequence of SEQ ID NO:29. The terminating oligonucleotide was made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:44 caguuucccaggcuuauccc, and the digestion oligonucleotide had the base sequence of SEQ ID NO:45 gtattagacccagtttcccaggct. The 5'-ends of the terminating oligonucleotide the hybridizing sequence of the promoter oligonucleotide identified in Example 2 overlapped by four bases, and the first 14 bases extending from the 5'-end of the digestion oligonucleotide overlapped with the 5'-most 14 bases of the hybridizing sequence of the promoter oligonucleotide. The blocked promoter oligonucleotide, the terminating oligonucleotide, and the digestion oligonucleotide all included a 3'-terminal blocking moiety consisting of the 3'-to-3' linkage described in Example 1. And the detection probe had the base sequence of SEQ ID NO:32 gctcatccca*caccgctaaagc, where the asterisk indicates the position of a standard AE label joined to the probe by means of a non-nucleotide linker. *See* Arnold *et al.*, U.S. Patent No. 5,585,481.

A total of six different reactions were performed in replicates of two, as set forth in Table 6 below. Template positive reactions were each provided with 1 µL of a 0.1% (w/v) LLS buffer containing 50 copies/µL of the *Mycobacterium tuberculosis* target nucleic acid of Example 2, and template negative reactions included no target nucleic acid. Amplification

and detection were essentially carried out as in Example 2 using 30 pmol/reaction each of the promoter and priming oligonucleotides, 5 pmol/reaction of the terminating oligonucleotide, 30 pmol/reaction of the digestion oligonucleotide, and 10 fmol/reaction of the detection probe. The results of these reactions, which were measured in relative light units ("RLU"), are presented in Table 6 and indicate that amplification in this amplification system was similar in the presence of either the terminating or the digestion oligonucleotide, although performance was somewhat better using the digestion oligonucleotide. Additionally, the results indicate that the level of amplification in this amplification system at this copy number was enhanced in the presence of the digestion oligonucleotide.

Table 6
Amplicon Production Using Terminating or Digestion Oligonucleotide

Reaction	Template	Terminating (T) or Digestion (D) Oligonucleotide	Promoter Oligonucleotide	RLU
1	Positive	T	Blocked	319,449
2		T	Blocked	254,181
3		T	Unblocked	20,915
4		T	Unblocked	3767
5		D	Blocked	472,786
6		D	Blocked	422,818
7		D	Unblocked	162,484
8		D	Unblocked	136,134
9		None	Blocked	10,007
10		None	Blocked	5052
11	Negative	D	Blocked	27,594
12		D	Blocked	5157

Example 9
Capped Priming Oligonucleotides

This experiment studied the effect of including a priming oligonucleotide cap on side-product formation using the *Mycobacterium tuberculosis* amplification system of Example 2. A "cap" is a short oligonucleotide complementary to the 3'-terminal end of a priming

oligonucleotide and includes a 3'-terminal blocking moiety to prevent extension from a terminal 3'-OH group. The cap is included to prevent the priming oligonucleotide from forming an oligonucleotide dimer with the promoter oligonucleotide, which could result in the formation of a functional double-stranded promoter sequence if the priming oligonucleotide is extended in the presence of a reverse transcriptase enzyme. As illustrated in FIG. 5A, the formation of an oligonucleotide dimer having a functional double-stranded promoter sequence could lead to the production of unwanted side-products in the presence of an RNA polymerase. While the cap inhibits oligonucleotide dimer formation, the cap can be readily displaced from the priming oligonucleotide through specific hybridization with the template sequence. A diagram of cap usage is shown in FIG. 6A.

For this experiment, we tested three different reaction conditions in replicates of two in the presence or absence of the *Mycobacterium tuberculosis* target nucleic acid of Example 2. The components of the three reaction conditions differed as follows: (i) the first set of reaction conditions included an unblocked promoter oligonucleotide, an uncapped priming oligonucleotide and a blocked terminating oligonucleotide; (ii) the second set of reaction conditions included a blocked promoter oligonucleotide, an uncapped priming oligonucleotide, and a blocked terminating oligonucleotide; and (iii) the third set of reaction conditions included a blocked promoter oligonucleotide, a priming oligonucleotide hybridized to a blocked cap at its 3'-terminal end, and a blocked terminating oligonucleotide. As in Example 2, the promoter oligonucleotide had the base sequence of SEQ ID NO:28 and the priming oligonucleotide had the base sequence of SEQ ID NO:29. The cap had the base sequence of SEQ ID NO:46 ctatc. The terminating oligonucleotide was made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:44 caguuucccaggcuuauccc. And the terminating oligonucleotide, the promoter oligonucleotide, when blocked, and the cap all included a 3'-terminal blocking moiety consisting of the 3'-to-3' linkage described in Example 1.

Prior to initiating amplification, the priming oligonucleotide and the cap were pre-hybridized in a 10 mM NaCl solution containing the priming oligonucleotide and the cap at a 1:1 ratio. The facilitate hybridization, the reaction tubes containing the solution were incubated in a 95°C water bath for 10 minutes and then cooled at room temperature for 2 hours. Following this pre-hybridization step, amplification was carried out as in Example 2 using 30 pmol/reaction each of the promoter oligonucleotide and the capped priming oligonucleotide and 5 pmol/reaction of the terminating oligonucleotide, where each reaction mixture was also provided with 1 µL of a 0.1% (w/v) LLS buffer containing 10,000 copies/µL of the target nucleic acid. After amplification, a 5 µL sample was taken from each tube, diluted 1:1 with a 10X BlueJuice™ Gel Loading Buffer (Invitrogen; Cat. No. 10816-015)

which was diluted to 2X with TBE (Tris-Borate-EDTA), and loaded onto an E-Gel® Single Comb Gel (4% high resolution agarose) which was pre-stained with ethidium bromide (Invitrogen; Cat. No. G5018-04). The gels were run on an E-Gel® Base (Invitrogen; Cat. No. G5100-01) at 80 volts for 30 minutes. The gels were then visualized on a FisherBiotech® Ultraviolet Transilluminator and photographed with a handheld camera using Polaroid 667 film .

The results of this experiment are illustrated in the photographed gels of FIG. 7A (template negative gel) and FIG. 7B (template positive gel). The numbers above the pictured gels indicate distinct lanes, where lane 7 is blank, lane 8 is a 100 base pair RNA ladder, lane 9 is a 20 base pair RNA ladder, and the remainder of the lanes contain products from the following reaction mixtures: (i) lanes 1 and 2 correspond to reaction mixtures containing the unblocked promoter oligonucleotide, the uncapped priming oligonucleotide, and the blocked terminating oligonucleotide; (ii) lanes 3 and 4 correspond to reaction mixtures containing the blocked promoter oligonucleotide, the uncapped priming oligonucleotide, and the terminating oligonucleotide; and (iii) lanes 5 and 6 correspond to reaction mixtures containing the blocked promoter oligonucleotide, the capped priming oligonucleotide, and the terminating oligonucleotide. The results clearly show that capping the priming oligonucleotide resulted in a further reduction in side-product formation (the side-products, which are oligonucleotide dimers in these reactions, are in the 20-mer to 60-mer range, whereas the amplicon would be greater than 100 bases in length).

Example 10
Looped Priming Oligonucleotides

In this experiment, the effect of looped priming oligonucleotides on amplification in the *Mycobacterium tuberculosis* amplification system of Example 2 was examined. Looped priming oligonucleotides are a variety of the priming oligonucleotides and caps evaluated in Example 9. A looped priming oligonucleotide includes a cap which is joined at its 3'-end to the 5'-end of the priming oligonucleotide by means of a non-nucleotide linker (*e.g.*, abasic nucleotides). One advantage of a looped priming oligonucleotide is that reassociation of the priming oligonucleotide and the cap, in the absence of the targeted template, is faster when the two oligonucleotides are maintained in close proximity to each other. Another advantage of a looped priming oligonucleotide is that the priming oligonucleotide and the cap can be generated in a single synthesis procedure, as opposed to the time intensive syntheses of separate priming and cap oligonucleotides.

Comparison was made between an uncapped priming oligonucleotide and looped priming oligonucleotides having caps of varying lengths. The promoter, priming and

terminating oligonucleotides were the same as those used in Example 9, and the detection probe was the same as detection probe I used in Example 2. The detection probe was provided to the reaction mixtures in both "cold" and "hot" forms, for the reasons described in Example 5, and the cold:hot probe ratio of each reaction mixture was 250:1. The looped priming oligonucleotides had the following sequences, where each "n" represents an abasic nucleotide (Glen Research; Cat. No. 10-1924-xx):

Looped Priming Oligonucleotide I (LPO I): SEQ ID NO:47 ctatttnngccgtcaccccaccaaca agctgatag;

Looped Priming Oligonucleotide II (LPO II): SEQ ID NO:48 ctatcnnnnngccgtcaccccaccaacaagctgatag;

Looped Priming Oligonucleotide III (LPO III): SEQ ID NO:49 ctatnnnnngccgtcaccccaccaacaagctgatag;

Looped Priming Oligonucleotide IV (LPO IV): SEQ ID NO:50 ctatcannnnngccgtcaccc caccaacaagctgatag;

Looped Priming Oligonucleotide V (LPO V): SEQ ID NO:51 ctatcnnnngccgtcaccccaccaacaagctgatag;

Looped Priming Oligonucleotide VI (LPO VI): SEQ ID NO:52 ctatcannnngccgtcacccc accaacaagctgatag; and

Looped Priming Oligonucleotide VII (LPO VII): SEQ ID NO:53 ctatcagcttgttggnnnnn gccgtcaccccaccaacaagctgatag.

A different reaction mixture was prepared for each priming oligonucleotide, and the reaction mixtures were tested in replicates of three using 1000 copies of the *Mycobacterium tuberculosis* target nucleic acid of Example 2 obtained from 0.1% (w/v) LLS buffer containing 1000 copies/μL of the target nucleic acid. The amplification and detection steps were carried out as in Example 2 using 30 pmol/reaction each of the promoter and priming oligonucleotides, 5 pmol/reaction of the terminating oligonucleotide, 10 fmol/reaction of the hot probe, and 2.5 pmol/reaction of the cold probe. Signal from the tubes was measured in relative light units ("RLU") and the average RLU values are presented in Table 7 below. The results indicate that the template can be amplified using a looped priming oligonucleotide, and that a looped priming oligonucleotide having four abasic groups and a five base cap is optimal for the *Mycobacterium tuberculosis* amplification system.

Table 7
Effect of Looped Priming Oligonucleotides on Amplification

Priming Oligonucleotide	Avg. RLU
Uncapped	430,060
LPO I	292,541
LPO II	260,559
LPO III	281,304
LPO IV	136,398
LPO V	372,119
LPO VI	171,382
LPO VII	20,045

Example 11
Comparison of Looped Priming Oligonucleotides and Caps

This experiment evaluated the ability of looped priming oligonucleotides to inhibit primer-dependent side-product formation. For this experiment, looped priming oligonucleotides LPO V and LPO VII of Example 10 were compared with an uncapped priming oligonucleotide and a priming oligonucleotide having a 14 base cap. The uncapped and capped priming oligonucleotides were the same as the uncapped priming oligonucleotide used in Example 10, and the cap had the base sequence of SEQ ID NO:54 ctatcagcttgttg (the cap and the priming oligonucleotide were pre-hybridized as in Example 9). The terminating oligonucleotide was the same as the terminating oligonucleotide used in Example 10, and the detection probe targeted the complement of the priming oligonucleotide and had the base sequence of SEQ ID NO:29 gccgtcacccc*accaacaagctgatag, where the asterisk indicates the position of a standard AE label joined to the probe by means of a non-nucleotide linker. *See* Arnold *et al.*, U.S. Patent No. 5,585,481. The detection probe was provided to the reaction mixtures in both "cold" and "hot" forms, for the reasons described in Example 5, and the cold:hot probe ratio of each reaction mixture was 4000:1. As with the promoter and terminating oligonucleotides, the cap had a 3'-terminal blocking moiety consisting of the 3' to 3' linkage described in Example 1.

The reaction mixtures were all template-free and tested in replicates of three, with a different set of reaction mixtures being prepared for each priming oligonucleotide. The amplification and detection steps were carried out as in Example 2 using 30 pmol/reaction

each of the promoter and priming oligonucleotides, 5 pmol/reaction of the terminating oligonucleotide, 20 fmol/reaction of the hot probe, and 80 pmol/reaction of the cold probe. Signal from the tubes was measured in relative light units ("RLU") and the averages of those RLU values are set forth in Table 9 below. The results indicate that the capped priming oligonucleotide inhibited primer-dependent side-product formation to a greater extent than did the looped priming oligonucleotides, although use of the looped priming oligonucleotides resulted in less primer-dependent side-product formation than when the uncapped priming oligonucleotide was used in this amplification system.

Table 8

Inhibition of Primer-Dependent Side-product Formation Using Looped Priming Oligonucleotides and Caps

Priming Oligonucleotide	Avg. RLU
Uncapped	2,246,565
LPO V	1,497,699
LPO VII	1,040,960
Capped	106,134

Example 12

Comparison of RNA Transcript Production in the Presence and Absence of Extender Oligonucleotides

This experiment examined the effect of extender oligonucleotides on amplicon production in amplification reaction mixtures containing a blocked promoter oligonucleotide. The extender oligonucleotides of this experiment were either blocked or unblocked and had the base sequence of SEQ ID NO:55 cctccaggaccccccctcccgggagagccata. A 3'-end blocked terminating oligonucleotide was included that was made up of 2'-O-methyl ribonucleotides having the base sequence of SEQ ID NO:56 auggcuagacgcuuucugcgugaaga. The target nucleic acid ("target"), priming oligonucleotide and promoter oligonucleotide were the same as those used in Example 1. The blocking moiety of each blocked oligonucleotide used in this experiment was a 3'-terminal blocking moiety consisting of the 3'-to-3' linkage described in Example 1. Cold and hot probes were used for detection of transcription products and had the sequence of SEQ ID NO:7. The hot probe of this experiment was identical to the first detection probe used in Example 5.

Six groups of amplification reaction mixtures were tested in replicates of four as follows: (i) no extender oligonucleotide and no target (group I); (ii) no extender

oligonucleotide and 100 copies of target (group II); (iii) blocked extender oligonucleotide and no target (group III); (iv) blocked extender oligonucleotide and 100 copies of target (group IV); (v) unblocked extender oligonucleotide and no target (group V); (iv) unblocked extender oligonucleotide and 100 copies of target (group VI). Reaction tubes from the six groups were set-up with 30 µL Amplification Reagent containing 6 pmol of the priming oligonucleotide, 4 pmol of the promoter oligonucleotide and 0.8 pmol of the terminating oligonucleotide. The reaction tubes of groups III and IV contained 4 pmol of the blocked extender oligonucleotide, and the reaction tubes of groups V and VI contained 4 pmol of the unblocked extender oligonucleotide. As indicated above, the reaction tubes of groups II, IV and VI further contained 100 copies of target, while those of groups I, III and V contained no target. The reaction mixtures were overlaid with 200 µL Oil Reagent, and the tubes were then sealed and vortexed for 10 seconds before being incubated in a 60°C water bath for 10 minutes. The tubes were then transferred to a 41.5°C water bath and incubated for 15 minutes before adding 10 µL Enzyme Reagent. After adding Enzyme Reagent, the tubes were again sealed and hand-shaken horizontally for 5 to 10 seconds to fully mix the components of the reaction mixtures. The tubes were returned to the 41.5°C water bath and incubated for an additional 60 minutes to permit amplification of the target sequence. Following amplification, the tubes were removed from the 41.5°C water bath and placed in an ice water bath for two minutes.

Detection of RNA transcription products was performed essentially as described in Example 2 (reaction tubes were vortexed rather than hand-shaken) using 100 fmol/reaction of the hot probe and 300 pmol/reaction of the cold probe. The averaged results are set forth in Table 9 in relative light units ("RLU") and demonstrate that the extender oligonucleotides of this experiment contributed to faster rates of amplification. The coefficient of variation values ("%CV") appearing in Table 9 for the different reaction conditions tested constitute the standard deviation of the replicates over the mean of the replicates as a percentage.

Table 9

Effect of Extender Oligonucleotides on Amplicon Production

Copy Number	Extender Oligonucleotide	Avg. RLU	%CV
0	None	4239	13
100	None	70,100	28
0	Blocked	4721	30
100	Blocked	337,964	12
0	Unblocked	13,324	76
100	Unblocked	869,861	12

Example 13

Amplification of a DNA Target Sequence

This experiment was conducted to determine the sensitivity of an amplification system according to the present invention for a target region contained within double-stranded DNA ("dsDNA"). The exemplary dsDNA used in this experiment was a cloned transcript derived from the orfX gene of a methicillin-resistant strain of *Staphylococcus aureus*. Included for amplification were a priming oligonucleotide, a displacer oligonucleotide, a terminating oligonucleotide and a promoter oligonucleotide. The priming oligonucleotide targeted a sequence contained within the 3'-end of the target region and had the base sequence of SEQ ID NO:35; the displacer oligonucleotide targeted a sequence present in a target nucleic acid containing the target region 5' to the target region and had the base sequence of SEQ ID NO:57 cuugctcaattaacacaacccgcatc, where the underlined bases were LNAs; the terminating oligonucleotide targeted a sequence present in the target nucleic acid 3' to the target region and had the base sequence of SEQ ID NO:58 gcgttggttcaauuc, where the underlined bases were LNAs (the terminating oligonucleotide was not fully comprised of LNAs to limit folding); and the promoter oligonucleotide had the base sequence of SEQ ID NO:34 aatttaatacgactcactatagggagaacgcatgacccaagggcaaagcgactttg, where the underlined portion was a T7 promoter sequence (SEQ ID NO:3) and the remainder was a hybridizing sequence (SEQ ID NO:33) that targeted the complement of a sequence contained within the 5'-end of the target region. The promoter oligonucleotide and the terminating oligonucleotide both included a 3'-terminal blocking moiety consisting of the 3'-to-3' linkage described in Example 1. Detection was carried out in real-time with a molecular torch detection probe having the base

sequence of SEQ ID NO:59 ccgucauuggcggaucagacgg. The base sequence of this probe was comprised of 2'-O-methylribonucleotides and nucleotides 17 and 18, reading 5'-to-3', were joined to each other by a 9-carbon linker. The detection probe was synthesized to include interacting DABCYL and FAM labels using 3'-DABCYL CPG (Prime Synthesis, Inc., Aston, PA; Cat. No. N-9756-10) and a 6-FAM phosphoramidite joined to the 5'-end (BioGenex Inc., San Ramon, CA; Cat. No. BGX-3008-01). The probe was also synthesized to include a 9-carbon linker positioned between nucleotides 17 and 18 reading 5'-to-3' (Glen Research Corporation, Sterling, VA; Cat. No. 10-1909-90). The synthesized probe was purified by polyacrylamide gel electrophoresis and reverse phase HPLC. The target binding portion of the probe consisted of SEQ ID NO:37.

The amplification reactions of this experiment were performed in replicates of six for each copy number and, for this, each of twelve test wells of a microtiter plate was provided with 30 µL of an amplification reagent comprising 44.1 mM HEPES, 2.82% (w/v) trehalose, 30.6 mM $MgCl_2$, 33 mM KCl, 0.3 % (v/v) ethanol, 0.1% (w/v) methyl paraben, 0.02% (w/v) propyl paraben, 9.41 mM rATP, 1.8 mM rCTP, 11.8 rGTP, 1.8 mM UTP, 0.47 mM dATP, 0.47 mM dCTP, 0.47 mM dGTP, 0.47 mM dTTP, 15 pmol priming oligonucleotide, 8 pmol displacer oligonucleotide, and 0.5 pmol terminating oligonucleotide at pH 7.7. Each well also contained 0, 10, 100, 1000 or 10,000 copies of the transcript. The plate was covered with a sealing card and incubated for 10 minutes at 95°C in the equivalent of a DNA Engine Opticon® 2 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA; Cat. No. CFD3220) to denature the double-stranded transcript, and then for 5 minutes at 42°C to promote hybridization of the priming oligonucleotide, the displacer oligonucleotide and the terminating oligonucleotide to the target nucleic acid. After incubating, the plate was removed from the detection system and 10 µL of an enzyme reagent was added to each test well to initiate amplification. The enzyme reagent comprised 58 mM HEPES, 3.03% (w/v) trehalose, 50 mM N-acetyl-L-cysteine, 1.04 mM EDTA, 120 mM KCl, 10% (w/v) TRITON® X-100, 20% (v/v) glycerol, 15 pmol promoter oligonucleotide, 8 pmol detection probe, 360 RTU/µL MMLV reverse transcriptase ("RT"), and 80 U/µL T7 RNA polymerase at pH 7.0, where one RTU of RT activity incorporates 1 nmol of dT into a substrate in 20 minutes at 37°C and one U of T7 RNA polymerase activity produces 5 fmol of RNA transcript in 20 minutes at 37°C. Following enzyme addition, the plate was placed on a Thermomixer R (Eppendorf North America, Inc., Westbury, NY; Cat. No. 5355 29716) pre-warmed to 44°C, covered with a sealing card, and agitated for 30 seconds at 1400 rpm. The plate was returned to the 42°C detection system, where an amplification reaction monitored in real-time was

carried out for 100 minutes, with fluorescent readings being taken every 24 seconds. Detection depended upon conformational changes in the probes hybridized to amplification products, thereby resulting in the emission of detectable, fluorescent signals.

The results of this experiment are illustrated graphically in Figure 8, where time in minutes is plotted on the x-axis and relative fluorescent signal is plotted on the y-axis. The data show that this amplification system had a sensitivity of at least 1000 copies of transcript. None of the control samples (0 transcript) exhibited detectable amplification.

Example 14

Comparison of Systems for Amplifying a DNA Target Sequence

A set of experiments was conducted to compare the sensitivities of various related amplification systems of the present invention for a target region contained within double-stranded DNA. Like Example 13, the exemplary target nucleic acid used in this experiment was a cloned transcript derived from the orfX gene of a methicillin-resistant strain of *Staphylococcus aureus*. The priming oligonucleotide used in these experiments had the base sequence of SEQ ID NO:36. The remainder of the oligonucleotides were the same as those of Example 13. Except for the first experiment, each of the experiments excluded one of the following components of Example 13: (i) a terminating oligonucleotide; (ii) a displacer oligonucleotide; (iii) a priming oligonucleotide; and (iv) heat to denature the double-stranded target. The experiments were run in replicates of twelve following the general protocol of Example 13. The sets of replicates contained either 0 or 1000 copies of the transcript.

The results of this experiment are illustrated graphically in Figures 9A-E, where, again, time in minutes is plotted on the x-axis and relative fluorescent signal is plotted on the y-axis.

The results of these experiments show the following: (i) all replicates containing the target were positive when the amplification system was not modified (*see* Figure 9A); (ii) all replicates containing target were positive when the terminating oligonucleotide was excluded from the reaction mixture (*see* Figure 9B); (iii) 6 out of 12 replicates containing the target were positive when the displacer oligonucleotide was excluded from the reaction mixture (*see* Figure 9C); (iv) 3 out of 12 replicates containing the target were positive when the priming oligonucleotide was excluded from the reaction mixture (*see* Figure 9D); and (v) 4 out of 12 replicates containing the target were positive when the double-stranded target was not exposed to heat-denaturing conditions (*i.e.*, 95°C heat step) prior to enzyme addition (*see* Figure 9E). None of the control samples (0 transcript) exhibited detectable amplification.

* * * * *

While the present invention has been described and shown in considerable detail with reference to certain preferred embodiments, those skilled in the art will readily appreciate other embodiments of the present invention. Accordingly, the present invention is deemed to include all modifications and variations encompassed within the spirit and scope of the following appended claims.

CLAIMS

1. A method of synthesizing multiple copies of a target sequence, said method comprising:

treating a target nucleic acid comprising a DNA target sequence with a priming oligonucleotide which hybridizes to the 3'-end of said target sequence such that a primer extension reaction can be initiated therefrom;

extending said priming oligonucleotide in a primer extension reaction with a DNA polymerase to give a first DNA primer extension product, at least a portion of said first primer extension product being complementary to said target sequence;

treating said first primer extension product with a promoter oligonucleotide comprising first and second regions, said first region comprising a base sequence which corresponds to a region at the 5'-end of said target sequence and which hybridizes to said first primer extension product to form a promoter oligonucleotide:first primer extension product hybrid, and said second region comprising a promoter for an RNA polymerase and situated 5' to said first region, wherein said promoter oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom;

transcribing from said promoter oligonucleotide:first primer extension product hybrid multiple first RNA products complementary to at least a portion of said first primer extension product using an RNA polymerase which recognizes said promoter and initiates transcription therefrom, wherein the base sequences of said first RNA products are substantially identical to the base sequence of said target sequence,

provided that if said first primer extension product has a defined 3'-end, then said method further comprises treating said target nucleic acid with a binding molecule which binds to said target nucleic acid adjacent to or near the 5'-end of said target sequence, and

further provided that said priming oligonucleotide does not include an RNA region which hybridizes to said target nucleic acid and which is selectively degraded by an enzyme activity when hybridized to said target nucleic acid.

2. The method of claim 1, wherein any oligonucleotide provided in said method which comprises a promoter for an RNA polymerase is modified to prevent the initiation of DNA synthesis therefrom.

3. The method of claim 1 or 2, wherein the activity of said DNA polymerase in said method is substantially limited to the formation of said first primer extension product.

4. The method of any one of claims 1 to 3, wherein said priming oligonucleotide consists of deoxynucleotides and/or analogs thereof.

5. The method of any one of claims 1 to 4, wherein said promoter oligonucleotide is modified to include a blocking moiety situated at its 3'-terminus.

6. The method of claim 5, wherein said blocking moiety of said promoter oligonucleotide comprises a substituent selected from the group consisting of: a modified nucleotide, a nucleotide or a nucleotide sequence having a 3'-to-5' orientation, a 3' alkyl group, a 3'2'-dideoxynucleotide, a 3' cordycepin, a 3' alkane-diol residue, a 3' non-nucleotide moiety, a nucleotide sequence non-complementary to said target sequence, a nucleic acid binding protein, and mixtures thereof.

7. The method of any one of claims 1 to 6, wherein said promoter oligonucleotide further comprises an insertion sequence positioned between or adjacent to said first and second regions, and wherein the presence of said insertion sequence in said promoter oligonucleotide enhances the rate at which said RNA products are formed.

8. The method of any one of claims 1 to 7 further comprising, prior to extending said priming oligonucleotide in a primer extension reaction, exposing a double-stranded complex comprising said target nucleic acid to conditions sufficient to denature said complex.

9. The method of claim 8, wherein said conditions include heat.

10. The method of any one of claims 1 to 7, wherein said target nucleic acid is in a double-stranded complex, and wherein said method does not comprise, prior to extending said priming oligonucleotide in a primer extension reaction, exposing said complex to conditions sufficient to denature said complex.

11. The method of any one of claims 1 to 10 further comprising:

treating said first RNA products with said priming oligonucleotide to form a priming oligonucleotide:first RNA product hybrid such that a primer extension reaction can be initiated from said priming oligonucleotide;

extending said priming oligonucleotide in a primer extension reaction with said DNA polymerase to give a second DNA primer extension product complementary to said first RNA product, said second primer extension product having a 3'-end which is complementary to the 5'-end of said first RNA product;

separating said second primer extension product from said first RNA product using an enzyme which selectively degrades said first RNA product;

treating said second primer extension product with said promoter oligonucleotide to form a promoter oligonucleotide:second primer extension product hybrid;

extending the 3'-end of said second primer extension product in said promoter oligonucleotide:second primer extension product hybrid to add a sequence complementary to said second region of said promoter oligonucleotide; and

transcribing from said promoter oligonucleotide:second primer extension product hybrid multiple second RNA products complementary to said second primer extension product using said RNA polymerase, wherein the base sequences of said second RNA products are substantially identical to the base sequence of said target sequence.

12. The method of claim 11, wherein said priming oligonucleotide is extended using a reverse transcriptase having an RNAse H activity.

13. The method of claim 11, wherein said enzyme has an RNAse H activity, and wherein said enzyme is other than a reverse transcriptase.

14. The method of any one of claims 1 to 13 further comprising:

treating said target nucleic acid with a displacer oligonucleotide which hybridizes to said target nucleic acid upstream from said priming oligonucleotide such that a primer extension reaction can be initiated therefrom; and

extending said displacer oligonucleotide in a primer extension reaction with said DNA polymerase to give a third DNA primer extension product that displaces said first primer extension product from said target nucleic acid.

15. The method of claim 14, wherein the activity of said DNA polymerase in said method is substantially limited to the formation of said first, second and/or third primer extension products.

16. The method of claim 14 or 15, wherein a 5'-region of said displacer oligonucleotide includes one or more modifications for increasing the binding affinity of said displacer oligonucleotide for said target nucleic acid, and wherein said modifications do not prevent said displacer oligonucleotide from being extended in a primer extension reaction.

17. The method of claim 16, wherein said modifications are spaced at least 15 bases from the 3'-terminus of said displacer oligonucleotide.

18. The method of claim 16 or 17, wherein said modifications are LNAs.

19. The method of any one of claims 14 to 18, wherein said displacer oligonucleotide hybridizes to said target nucleic acid such that the 3'-terminal base of said displacer oligonucleotide is adjacent to the 5'-terminal base of said priming oligonucleotide.

20. The method of any one of claims 14 to18, wherein said displacer oligonucleotide hybridizes to said target nucleic acid such that the 3'-terminal base of said displacer oligonucleotide is spaced from 5 to 35 bases from the 5'-terminal base of said priming oligonucleotide.

21. The method of any one of claims 14 to 20, wherein said target nucleic acid is treated with said priming oligonucleotide prior to treating said target nucleic acid with said displacer oligonucleotide.

22. The method of any one of claims 14 to 21, wherein said first primer extension product is formed prior to treating said target nucleic acid with said displacer oligonucleotide.

23. The method of claim 14 to 21, wherein said target nucleic acid is treated with said priming oligonucleotide and said displacer oligonucleotide prior to exposing said target nucleic acid to a DNA polymerase.

24. The method of any one of claims 14 to 23 further comprising treating the 3'-end of at least one of said priming oligonucleotide and said displacer oligonucleotide with one or more caps, each said cap comprising a base region complementary to at least 3 bases at the

3'-end of said priming oligonucleotide or said displacer oligonucleotide, wherein the 5'-terminal base of each said cap is complementary to the 3'-terminal base of said priming oligonucleotide or said displacer oligonucleotide, and wherein each said cap is modified to prevent the initiation of DNA synthesis therefrom.

25. The method of claim 24, wherein each said cap is complementary to no more than 8 bases at the 3'-end of said priming oligonucleotide or said displacer oligonucleotide.

26. The method of claim 24 or 25, wherein each said cap prevents non-specific hybridization between said priming oligonucleotide or said displacer oligonucleotide and said promoter oligonucleotide when each said cap is hybridized to said priming oligonucleotide or said displacer oligonucleotide.

27. The method of any one of claims 24 to 26, wherein each said cap is a capping oligonucleotide modified to include a blocking moiety situated at its 3'-terminus.

28. The method of any one of claims 24 to 27, wherein the 3'-end of at least one of said caps is covalently attached to the 5'-end of said priming oligonucleotide or said displacer oligonucleotide, and wherein said at least one cap hybridizes to the 3'-end of said priming oligonucleotide or said displacer oligonucleotide by forming a loop.

29. The method of claim 28, wherein said at least one cap is joined to said priming oligonucleotide or said displacer oligonucleotide via a linker region.

30. The method of claim 29, wherein said linker region comprises at least 5 nucleotides.

31. The method of claim 29, wherein said linker region comprises at least 5 abasic nucleotides.

32. The method of any one of claims 1 to 31 further comprising:

treating said target nucleic acid with said binding molecule; and

extending the 3'-end of said first primer extension product in said promoter oligonucleotide:first primer extension product hybrid to add a sequence complementary to said second region of said promoter oligonucleotide.

33. The method of claim 32, wherein said binding molecule comprises an oligonucleotide having a blocking moiety situated at its 3'-terminus to prevent the initiation of DNA synthesis therefrom.

34. The method of claim 32 or 33, wherein said binding molecule comprises LNAs.

35. The method of any one of claims 32 to 34, wherein said binding molecule is a terminating oligonucleotide.

36. The method of any one of claims 32 to 35, wherein the 5'-end of said terminating oligonucleotide is complementary to at least two bases at the 5'-end of said first region of said promoter oligonucleotide.

37. The method of any one of claims 32 to 35, wherein the 5'-end of said terminating oligonucleotide is complementary to at least three but no more than ten bases at the 5'-end of said first region of said promoter oligonucleotide.

38. The method of any one of claims 32 to 34, wherein said binding molecule is a modifying oligonucleotide.

39. The method of claim 38, wherein said modifying oligonucleotide is a digestion oligonucleotide.

40. The method of any one of claims 1 to 39 further comprising treating said first primer extension product with an extender oligonucleotide, said extender oligonucleotide hybridizing to a region of said first primer extension product 3' to said promoter oligonucleotide of said promoter oligonucleotide:first primer extension product hybrid, such that an extender oligonucleotide:first primer extension product hybrid is formed.

41. The method of claim 40, wherein said extender oligonucleotide further comprises a blocking moiety situated at its 3' terminus to prevent the initiation of DNA synthesis therefrom.

42. The method of claim 40 or 41, wherein said extender oligonucleotide hybridizes to said first primer extension product such that the 5'-terminal base of said extender oligonucleotide is spaced within three nucleotides of the 3'-terminal base of said promoter oligonucleotide.

43. The method of claim 40 or 41, wherein said extender oligonucleotide hybridizes to said first primer extension product adjacent said promoter oligonucleotide.

44. The method of any one of claims 1 to 43 further comprising determining the presence or amount of said multiple copies of said target sequence.

45. The method of claim 44, wherein the presence or amount of said multiple copies of said target sequence is determined with a detection probe having a detectable label.

46. The method of claim 45, wherein the presence or amount of said multiple copies of said target sequence is determined after said first RNA products are transcribed.

47. The method of claim 45, wherein the presence or amount of said multiple copies of said target sequence is determined as said first RNA products are being transcribed.

48. The method of claim 47, wherein said probe is a self-hybridizing probe and includes a pair of interacting labels.

49. The method of any one of claims 1 to 48, wherein said method is carried out at a substantially constant temperature.

50. A kit for use in amplifying a DNA target sequence, said kit comprising:

a priming oligonucleotide which hybridizes to the 3'-end of a DNA target sequence such that a primer extension reaction can be initiated therefrom; and

a promoter oligonucleotide comprising first and second regions, wherein said first region has a base sequence which corresponds to a region at the 5'-end of said target sequence and which hybridizes to a 3'-region of a DNA primer extension product comprising said priming oligonucleotide, wherein said second region is a promoter for an RNA polymerase and is situated 5' to said first region, and wherein said promoter oligonucleotide is modified to prevent the initiation of DNA synthesis therefrom,

provided that any oligonucleotide included in said kit that comprises a promoter for an RNA polymerase is modified to prevent the initiation of DNA synthesis therefrom,

provided that said kit does not include a restriction endonuclease, and

further provided that said priming oligonucleotide does not include an RNA region which hybridizes to said target nucleic acid and which is selectively degraded by an enzyme activity when hybridized to said target nucleic acid.

51. The kit of claim 50, wherein said priming oligonucleotide consists of deoxynucleotides and/or analogs thereof.

52. The kit of claim 50 or 51, wherein said promoter oligonucleotide is modified to include a blocking moiety situated at its 3'-terminus.

53. The kit of claim 52, wherein said blocking moiety of said promoter oligonucleotide comprises a substituent selected from the group consisting of: a modified nucleotide, a nucleotide or a nucleotide sequence having a 3'-to-5' orientation, a 3' alkyl group, a 3'2'-dideoxynucleotide, a 3' cordycepin, a 3' alkane-diol residue, a 3' non-nucleotide moiety, a nucleotide sequence non-complementary to said target sequence, a nucleic acid binding protein, and mixtures thereof.

54. The kit of any one of claims 50 to 53, wherein said promoter oligonucleotide further comprises an insertion sequence positioned between or adjacent to said first and second regions, and wherein the presence of said insertion sequence in said promoter oligonucleotide enhances the rate at which said RNA products are formed.

55. The kit of any one of claims 50 to 54 further comprising a chemical denaturant for rendering single-stranded a double-stranded DNA complex.

56. The kit of any one of claims 50 to 55 further comprising a reverse transcriptase having an RNAse H activity, wherein said kit does not include an enzyme other than said reverse transcriptase having an RNAse H activity.

57. The kit of any one of claims 50 to 55 further comprising an enzyme, other than a reverse transcriptase, having an RNAse H activity.

58. The kit of any one of claims 50 to 57 further comprising a displacer oligonucleotide which hybridizes to a target nucleic acid containing said target sequence upstream from said priming oligonucleotide such that a primer extension reaction can be initiated therefrom.

59. The kit of claim 58, wherein a 5'-region of said displacer oligonucleotide includes one or more modifications for increasing the binding affinity of said displacer oligonucleotide for said target nucleic acid, and wherein said modifications do not prevent said displacer oligonucleotide from being extended in a primer extension reaction.

60. The kit of claim 59, wherein said modifications are spaced at least 15 bases from the 3'-terminus of said displacer oligonucleotide.

61. The kit of claim 59 or 60, wherein said modifications are LNAs.

62. The kit of any one of claims 58 to 61, wherein said displacer oligonucleotide hybridizes to said target nucleic acid such that the 3'-terminal base of said displacer oligonucleotide is adjacent to the 5'-terminal base of said priming oligonucleotide.

63. The kit of any one of claims 58 to 61, wherein said displacer oligonucleotide hybridizes to said target nucleic acid such that the 3'-terminal base of said displacer oligonucleotide is spaced from 5 to 35 bases from the 5'-terminal base of said priming oligonucleotide.

64. The kit of any one of claims 58 to 63, wherein at least one of said priming oligonucleotide and said displacer oligonucleotide includes a cap hybridized to the 3'-end thereof, said cap comprising a base region complementary to at least 3 bases at the 3'-end of said priming oligonucleotide or said displacer oligonucleotide, wherein the 5'-terminal base of said cap is complementary to the 3'-terminal base of said priming oligonucleotide or said displacer oligonucleotide, and wherein said cap is modified to prevent the initiation of DNA synthesis therefrom.

65. The kit of claim 64, wherein said cap is complementary to no more than 8 bases at the 3'-end of said priming oligonucleotide or said displacer oligonucleotide.

66. The kit of claim 64 or 65, wherein said cap prevents non-specific hybridization between said priming oligonucleotide or said displacer oligonucleotide and said promoter oligonucleotide when each said cap is hybridized to said priming oligonucleotide or said displacer oligonucleotide.

67. The kit of any one of claims 64 to 66, wherein said cap is a capping oligonucleotide modified to include a blocking moiety situated at its 3'-terminus.

68. The kit of any one of claims 64 to 67, wherein the 3'-end of said cap is covalently attached to the 5'-end of said priming oligonucleotide or said displacer oligonucleotide, and wherein said cap hybridizes to the 3'-end of said priming oligonucleotide or said displacer oligonucleotide by forming a loop.

69. The kit of claim 68, wherein said cap is joined to said priming oligonucleotide or said displacer oligonucleotide via a linker region.

70. The kit of claim 69, wherein said linker region comprises at least 5 nucleotides.

71. The kit of claim 69, wherein said linker region comprises at least 5 abasic nucleotides.

72. The kit of claim 50 to 71 further comprising a binding molecule which binds to a target nucleic acid containing said target sequence adjacent to or near the 5'-end of said target sequence.

73. The kit of claim 72, wherein said binding molecule comprises an oligonucleotide having a blocking moiety situated at its 3'-terminus to prevent the initiation of DNA synthesis therefrom.

74. The kit of claim 72 or 73, wherein said binding molecule comprises LNAs.

75. The kit of any one of claims 72 to 74, wherein said binding molecule is a terminating oligonucleotide.

76. The kit of claim 75, wherein the 5'-end of said terminating oligonucleotide is complementary to at least two bases at the 5'-end of said first region of said promoter oligonucleotide.

77. The kit of claim 75, wherein the 5'-end of said terminating oligonucleotide is complementary to at least three but no more than ten bases at the 5'-end of said first region of said promoter oligonucleotide.

78. The kit of any one of claims 72 to 74, wherein said binding molecule is a modifying oligonucleotide.

79. The kit of claim 78, wherein said modifying oligonucleotide is a digestion oligonucleotide.

80. The kit of any one of claims 50 to 79 further comprising an extender oligonucleotide which hybridizes to a region of said primer extension product 3' to said promoter oligonucleotide of said promoter oligonucleotide:primer extension product hybrid, such that an extender oligonucleotide:primer extension product hybrid is formed.

81. The kit of claim 80, wherein said extender oligonucleotide further comprises a blocking moiety situated at its 3' terminus to prevent the initiation of DNA synthesis therefrom.

82. The kit of claim 80 or 81, wherein said extender oligonucleotide hybridizes to said primer extension product such that the 5'-terminal base of said extender oligonucleotide is spaced within three nucleotides of the 3'-terminal base of said promoter oligonucleotide.

83. The kit of claim 80 or 81, wherein said extender oligonucleotide hybridizes to said primer extension product adjacent said promoter oligonucleotide.

84. The kit of claim 50 to 83 further comprising a detection probe which hybridizes to said target sequence.

85. The kit of claim 84, wherein said detection probe includes a detectable label.

86. The kit of claim 84, wherein said probe is a self-hybridizing probe having a pair of interacting labels.

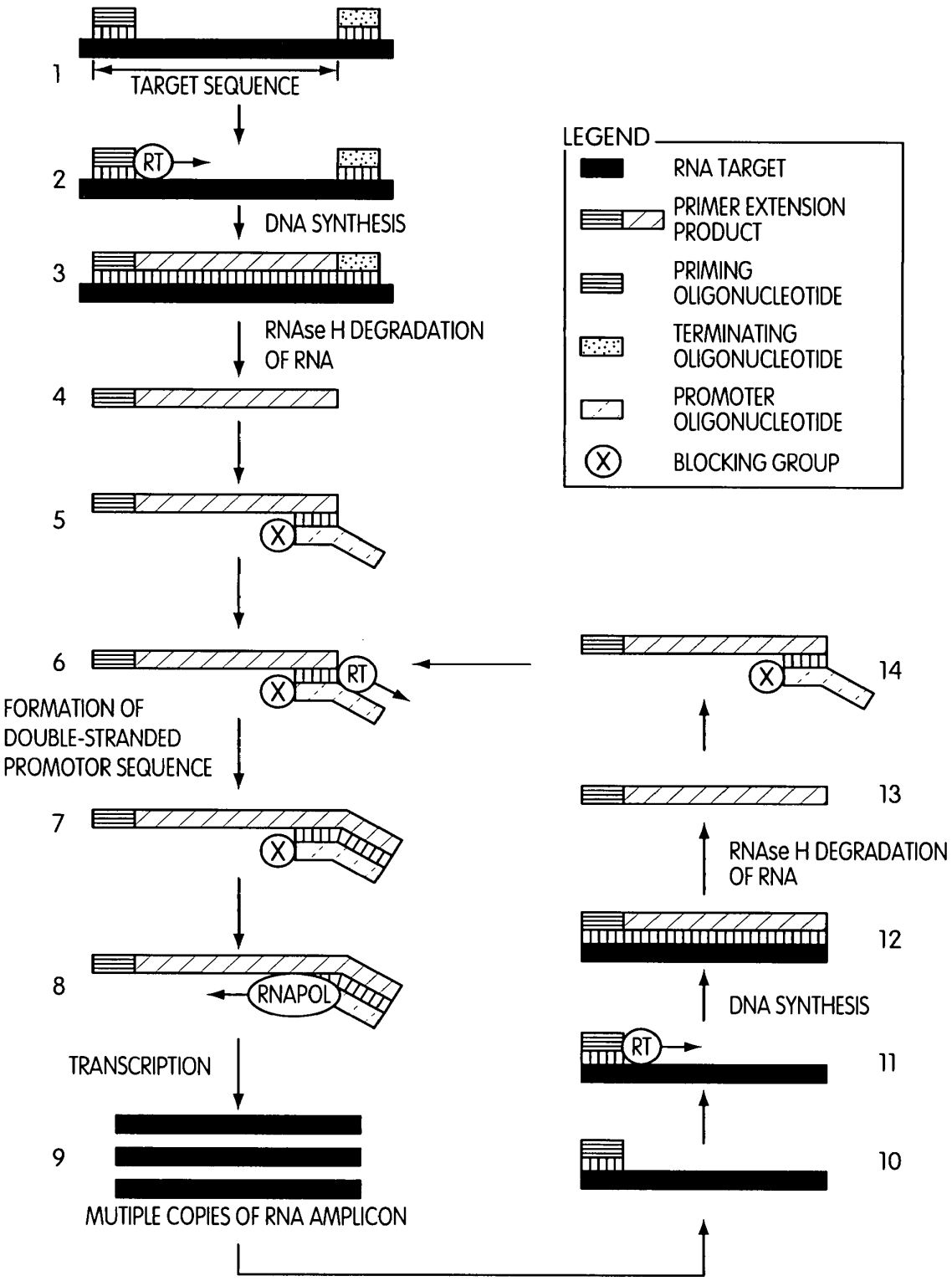

1
TARGET SEQUENCE
2
RT
DNA SYNTHESIS
3
RNAse H DEGRADATION OF RNA
4
5
X
6
RT
X
FORMATION OF DOUBLE-STRANDED PROMOTOR SEQUENCE
7
X
8
RNAPOL
TRANSCRIPTION
9
MUTIPLE COPIES OF RNA AMPLICON
LEGEND
RNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
TERMINATING OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
BLOCKING GROUP
14
13
RNAse H DEGRADATION OF RNA
12
DNA SYNTHESIS
RT
11
10

FIG. 1A

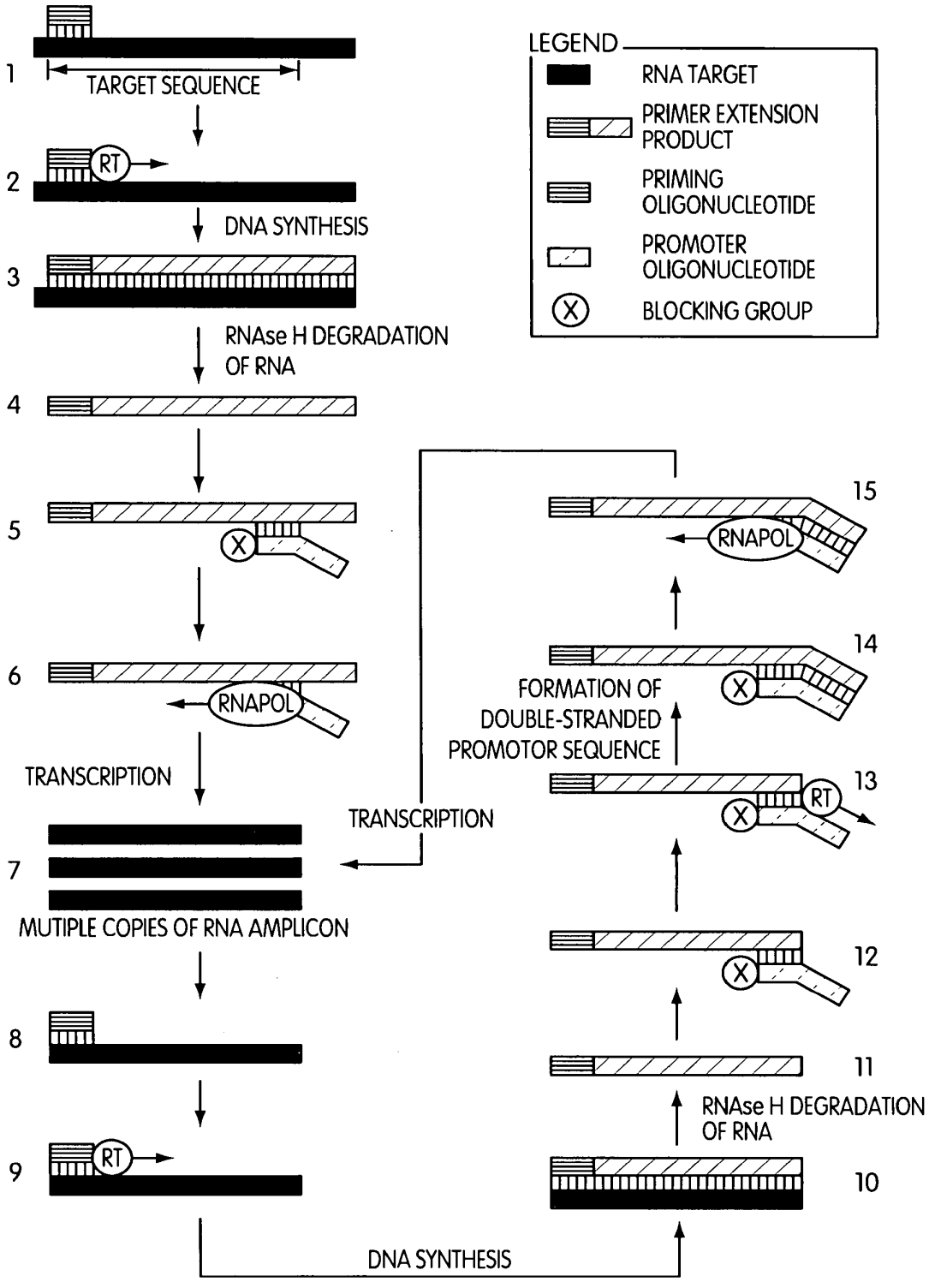

LEGEND
RNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
BLOCKING GROUP
1
TARGET SEQUENCE
2
RT
DNA SYNTHESIS
3
RNAse H DEGRADATION OF RNA
4
5
X
6
RNAPOL
TRANSCRIPTION
7
MUTIPLE COPIES OF RNA AMPLICON
8
9
RT
DNA SYNTHESIS
10
RNAse H DEGRADATION OF RNA
11
12
13
RT
FORMATION OF DOUBLE-STRANDED PROMOTOR SEQUENCE
14
15
RNAPOL
TRANSCRIPTION

FIG. 1B

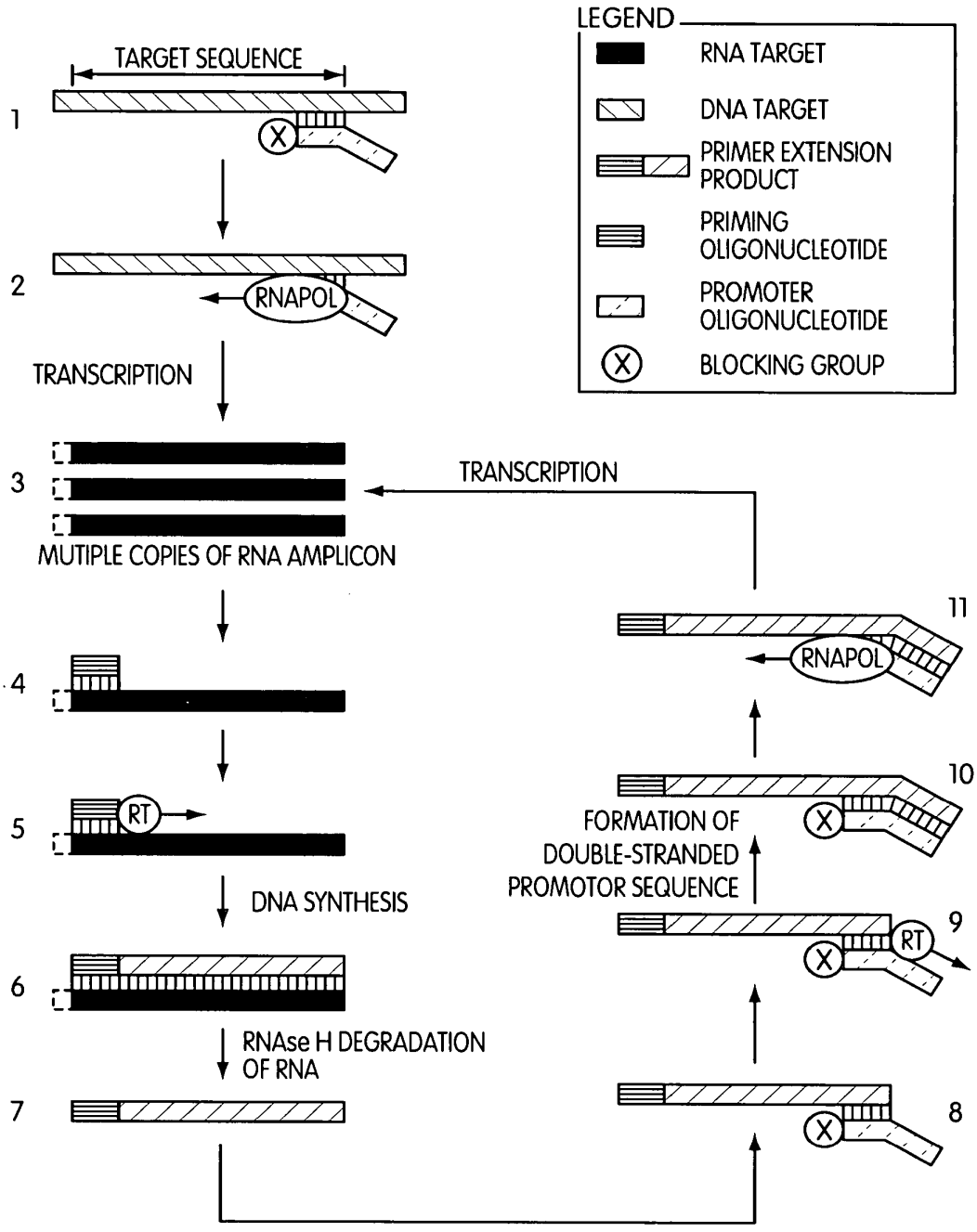

LEGEND
RNA TARGET
DNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
BLOCKING GROUP
TARGET SEQUENCE
1
2
RNAPOL
TRANSCRIPTION
3
MUTIPLE COPIES OF RNA AMPLICON
TRANSCRIPTION
4
5
RT
DNA SYNTHESIS
6
RNAse H DEGRADATION OF RNA
7
8
9
RT
FORMATION OF DOUBLE-STRANDED PROMOTOR SEQUENCE
10
11
RNAPOL

FIG. 1C

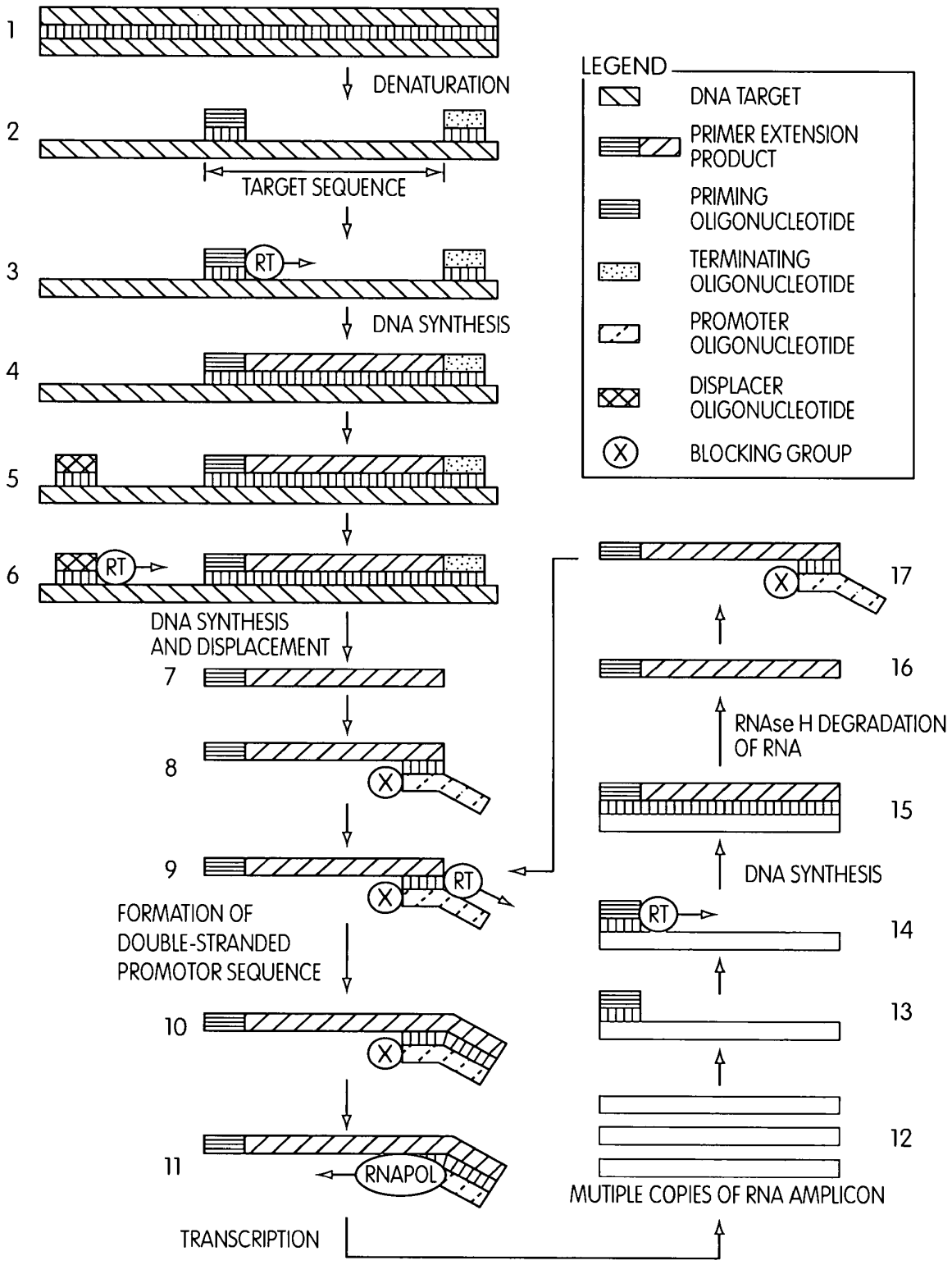

1
DENATURATION
2
TARGET SEQUENCE
3
RT
DNA SYNTHESIS
4
5
6
RT
DNA SYNTHESIS
AND DISPLACEMENT
7
8
X
9
RT
X
FORMATION OF
DOUBLE-STRANDED
PROMOTOR SEQUENCE
10
X
11
RNAPOL
TRANSCRIPTION
LEGEND
DNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
TERMINATING OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
DISPLACER OLIGONUCLEOTIDE
X
BLOCKING GROUP
17
X
16
RNAse H DEGRADATION
OF RNA
15
DNA SYNTHESIS
14
RT
13
12
MUTIPLE COPIES OF RNA AMPLICON

FIG. 1D

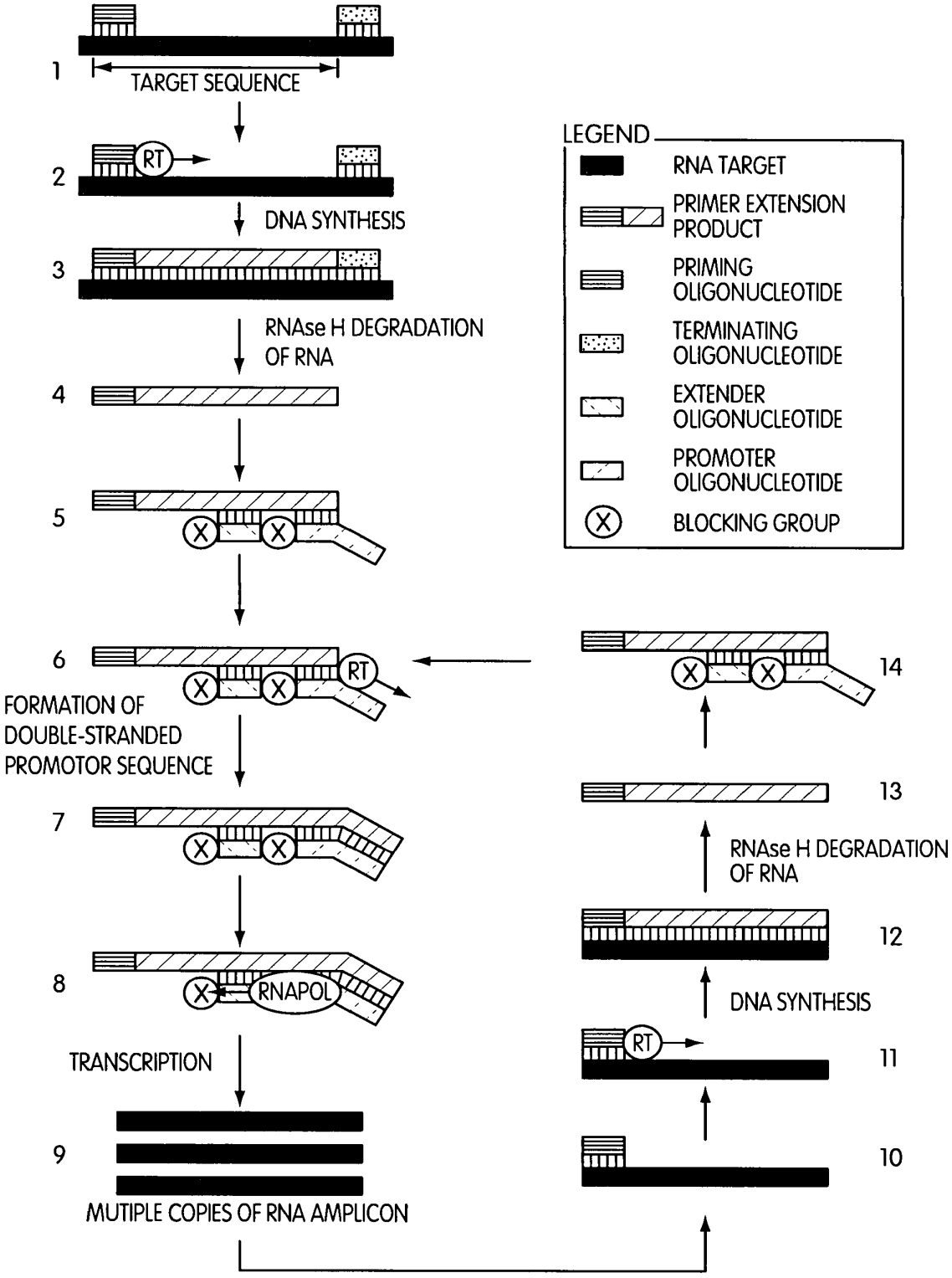

1
TARGET SEQUENCE
2
RT
DNA SYNTHESIS
3
RNAse H DEGRADATION OF RNA
4
5
6
RT
FORMATION OF DOUBLE-STRANDED PROMOTOR SEQUENCE
7
8
RNAPOL
TRANSCRIPTION
9
MUTIPLE COPIES OF RNA AMPLICON
LEGEND
RNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
TERMINATING OLIGONUCLEOTIDE
EXTENDER OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
BLOCKING GROUP
14
13
RNAse H DEGRADATION OF RNA
12
DNA SYNTHESIS
RT
11
10

FIG. 2A

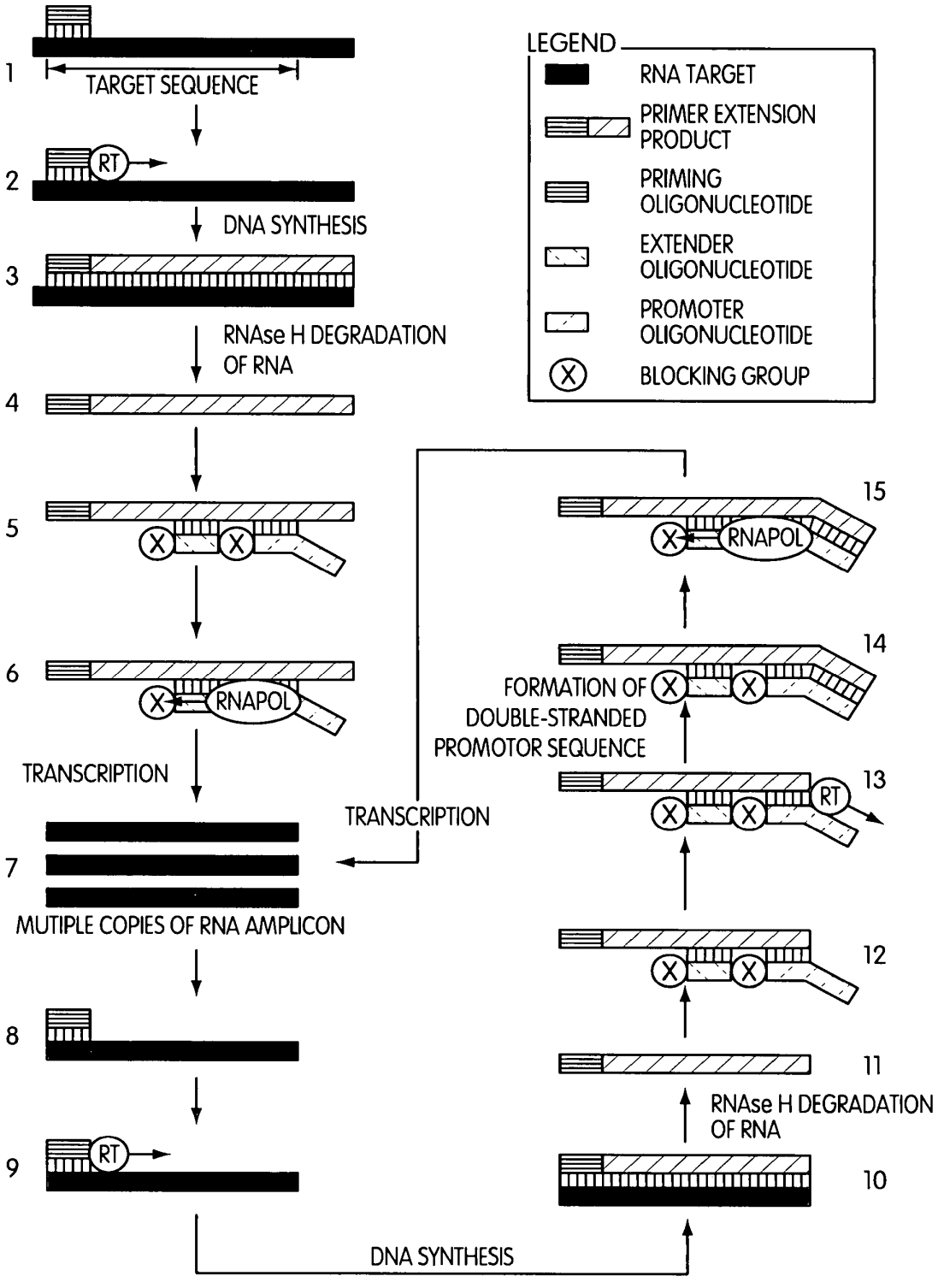

LEGEND
RNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
EXTENDER OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
BLOCKING GROUP
1
TARGET SEQUENCE
2
RT
DNA SYNTHESIS
3
RNAse H DEGRADATION OF RNA
4
5
6
RNAPOL
TRANSCRIPTION
7
MUTIPLE COPIES OF RNA AMPLICON
8
9
RT
DNA SYNTHESIS
10
RNAse H DEGRADATION OF RNA
11
12
13
RT
FORMATION OF DOUBLE-STRANDED PROMOTOR SEQUENCE
14
15
RNAPOL
TRANSCRIPTION

FIG. 2B

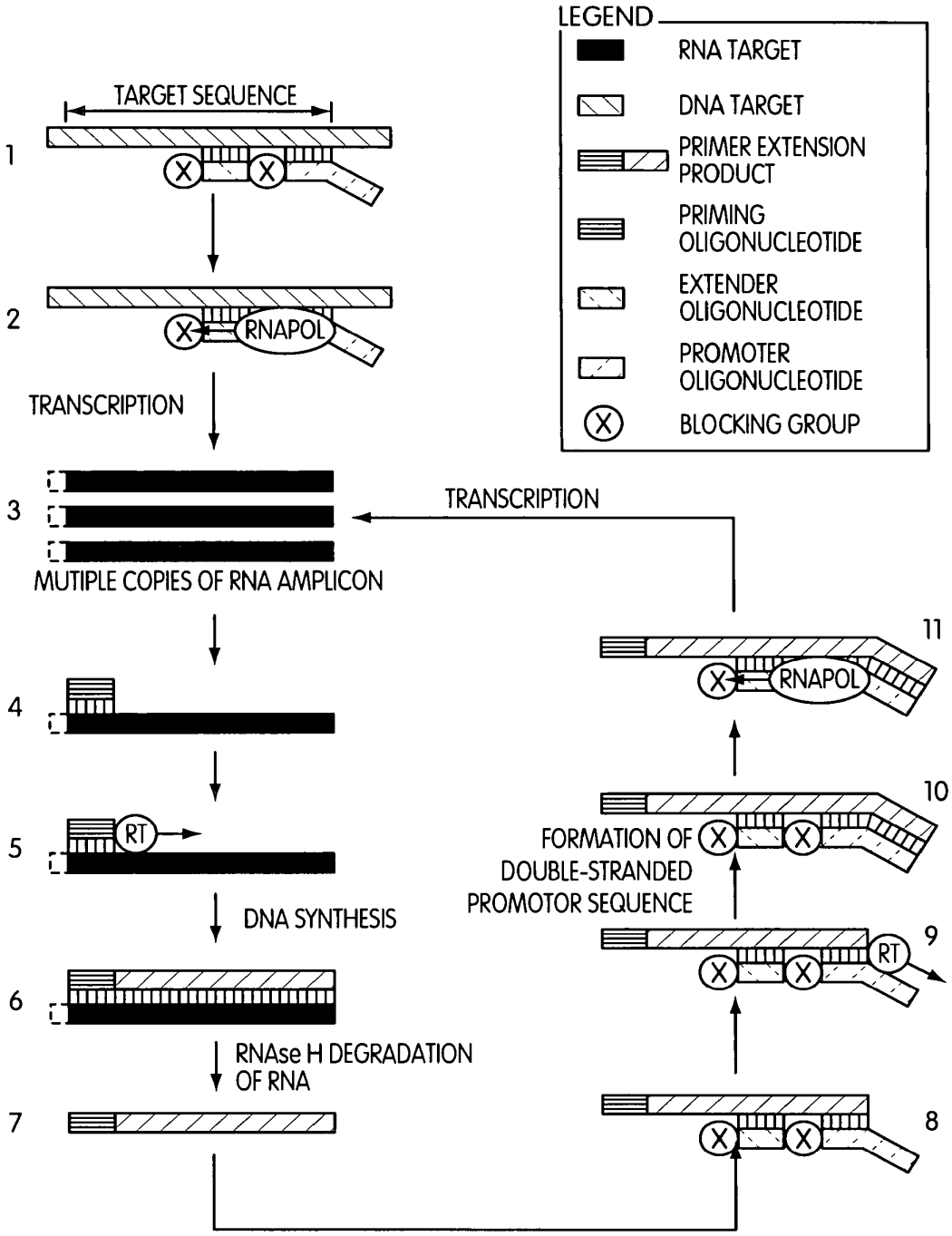

LEGEND
RNA TARGET
DNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
EXTENDER OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
BLOCKING GROUP
TARGET SEQUENCE
1
2
RNAPOL
TRANSCRIPTION
3
MUTIPLE COPIES OF RNA AMPLICON
TRANSCRIPTION
4
5
RT
DNA SYNTHESIS
6
RNAse H DEGRADATION OF RNA
7
8
9
RT
10
FORMATION OF DOUBLE-STRANDED PROMOTOR SEQUENCE
11
RNAPOL

FIG. 2C

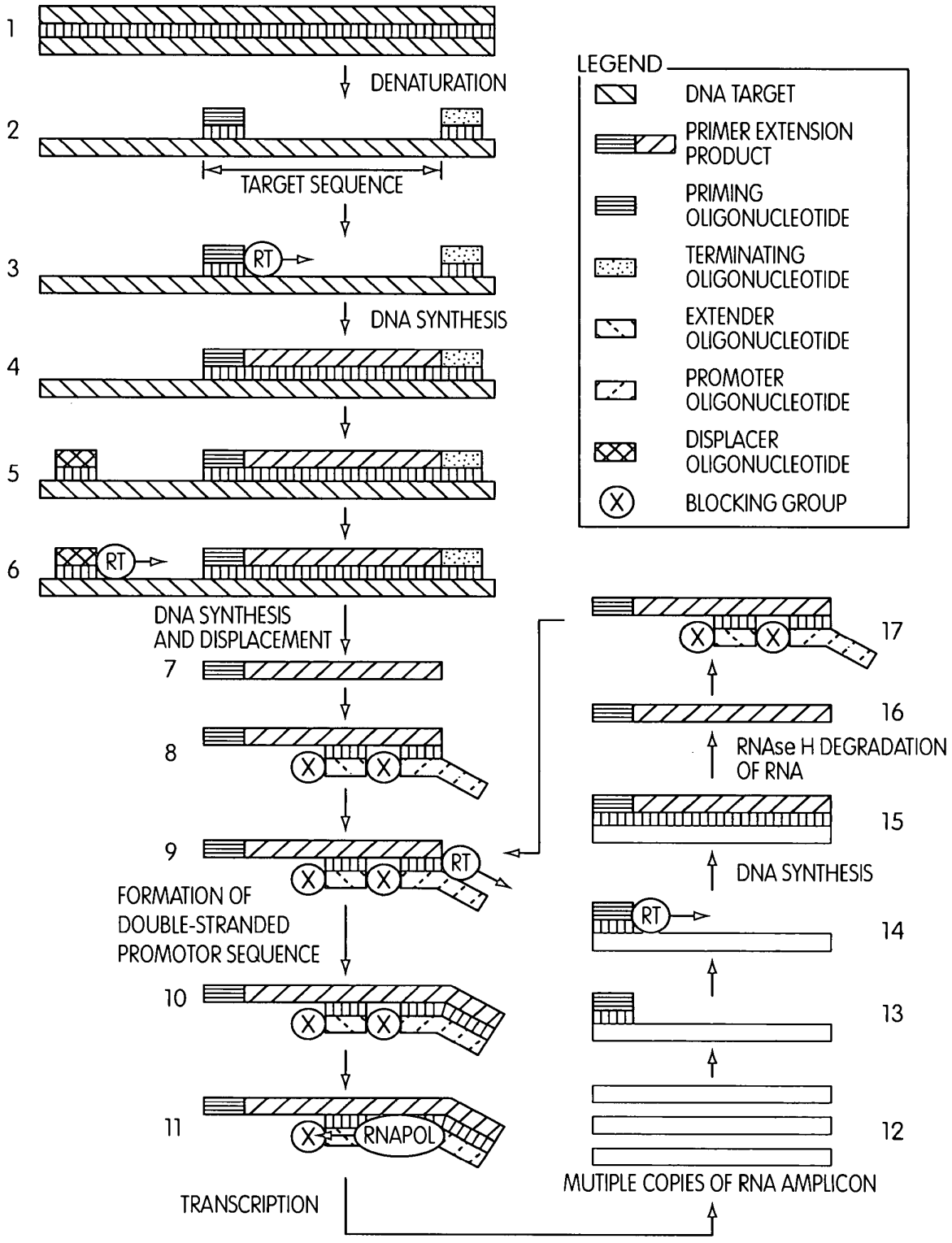

1
DENATURATION
2
TARGET SEQUENCE
3
RT
DNA SYNTHESIS
4
5
6
RT
DNA SYNTHESIS
AND DISPLACEMENT
7
8
9
RT
FORMATION OF
DOUBLE-STRANDED
PROMOTOR SEQUENCE
10
11
RNAPOL
TRANSCRIPTION
LEGEND
DNA TARGET
PRIMER EXTENSION PRODUCT
PRIMING OLIGONUCLEOTIDE
TERMINATING OLIGONUCLEOTIDE
EXTENDER OLIGONUCLEOTIDE
PROMOTER OLIGONUCLEOTIDE
DISPLACER OLIGONUCLEOTIDE
BLOCKING GROUP
17
16
RNAse H DEGRADATION
OF RNA
15
DNA SYNTHESIS
RT
14
13
12
MUTIPLE COPIES OF RNA AMPLICON

FIG. 2D

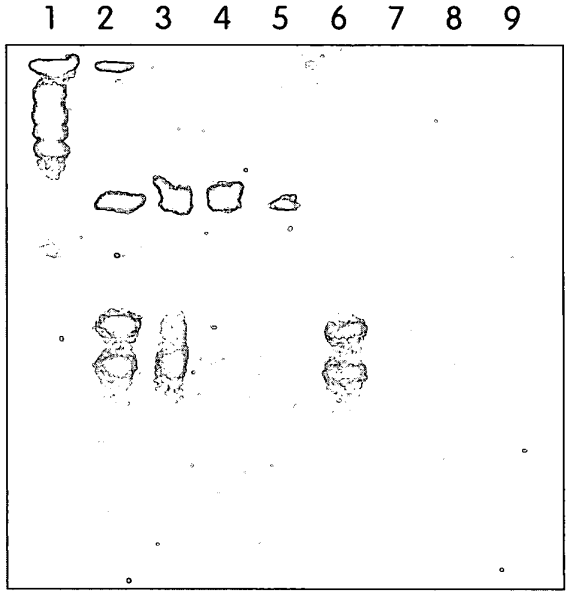


FIG. 3

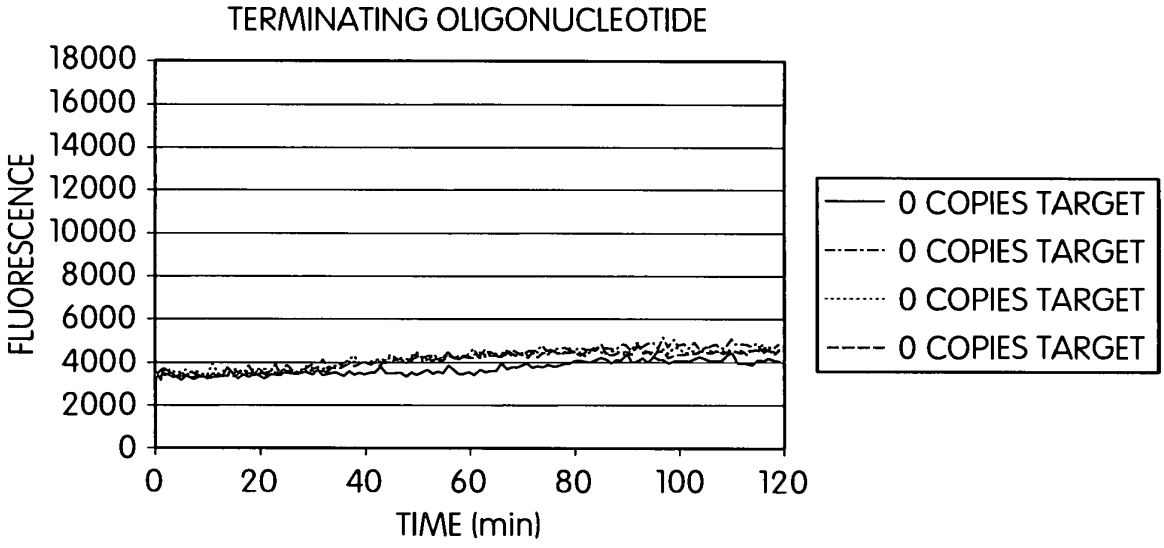

TERMINATING OLIGONUCLEOTIDE
FLUORESCENCE
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
TIME (min)
0 COPIES TARGET
0 COPIES TARGET
0 COPIES TARGET
0 COPIES TARGET

FIG. 4A

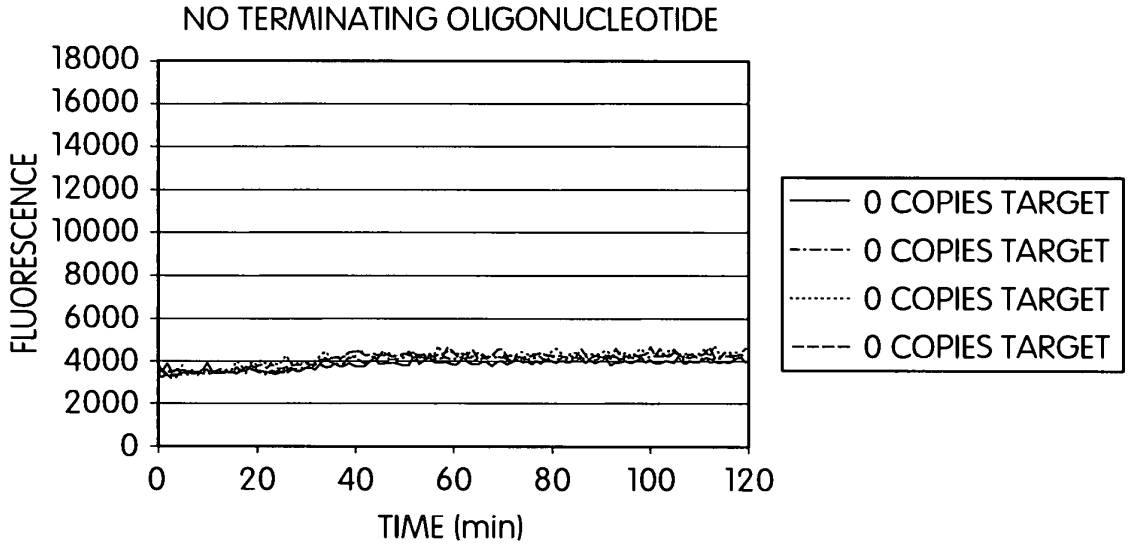

NO TERMINATING OLIGONUCLEOTIDE
FLUORESCENCE
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
TIME (min)
0 COPIES TARGET
0 COPIES TARGET
0 COPIES TARGET
0 COPIES TARGET

FIG. 4B

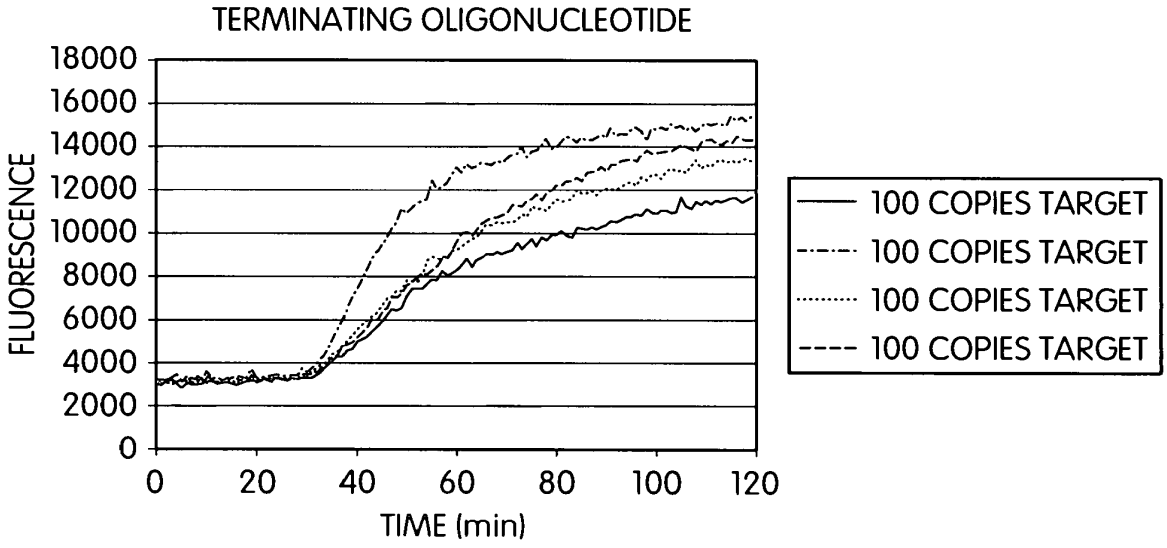

TERMINATING OLIGONUCLEOTIDE
FLUORESCENCE
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
TIME (min)
100 COPIES TARGET
100 COPIES TARGET
100 COPIES TARGET
100 COPIES TARGET

FIG. 4C

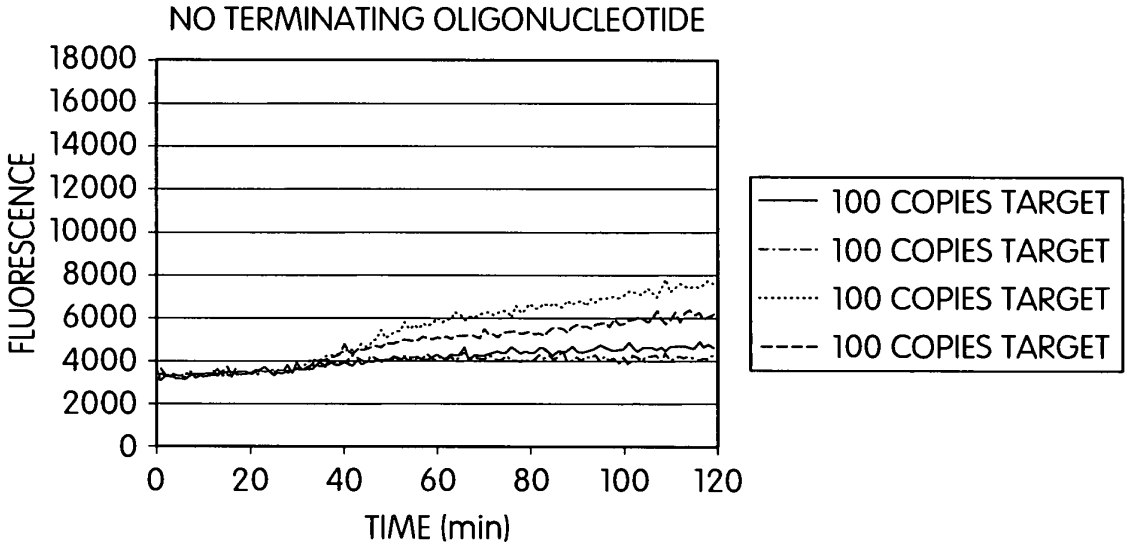

NO TERMINATING OLIGONUCLEOTIDE
FLUORESCENCE
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
TIME (min)
100 COPIES TARGET
100 COPIES TARGET
100 COPIES TARGET
100 COPIES TARGET

FIG. 4D

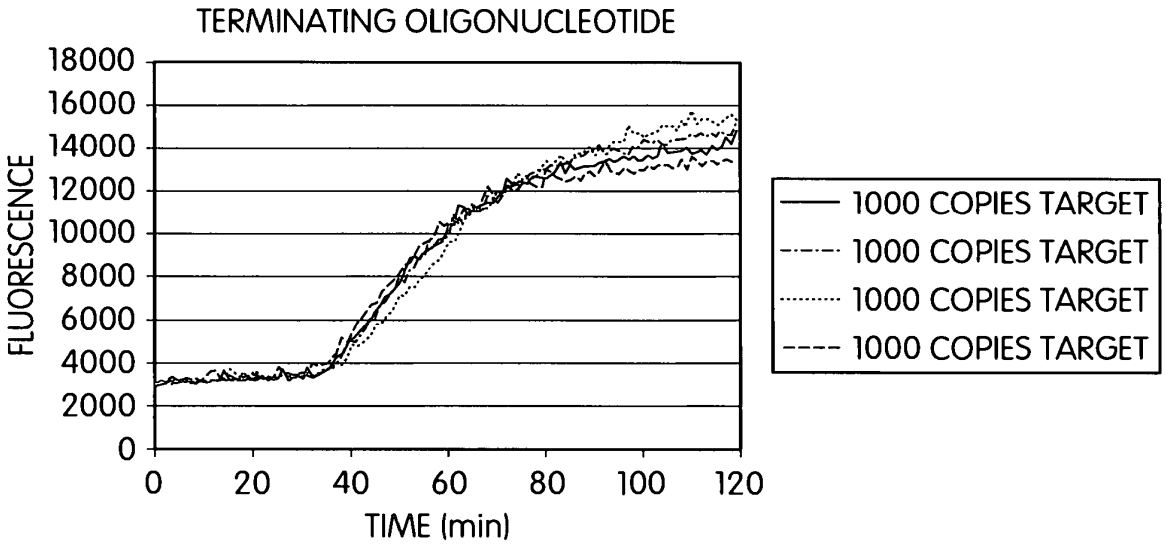

TERMINATING OLIGONUCLEOTIDE
FLUORESCENCE
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
TIME (min)
1000 COPIES TARGET
1000 COPIES TARGET
1000 COPIES TARGET
1000 COPIES TARGET

FIG. 4E

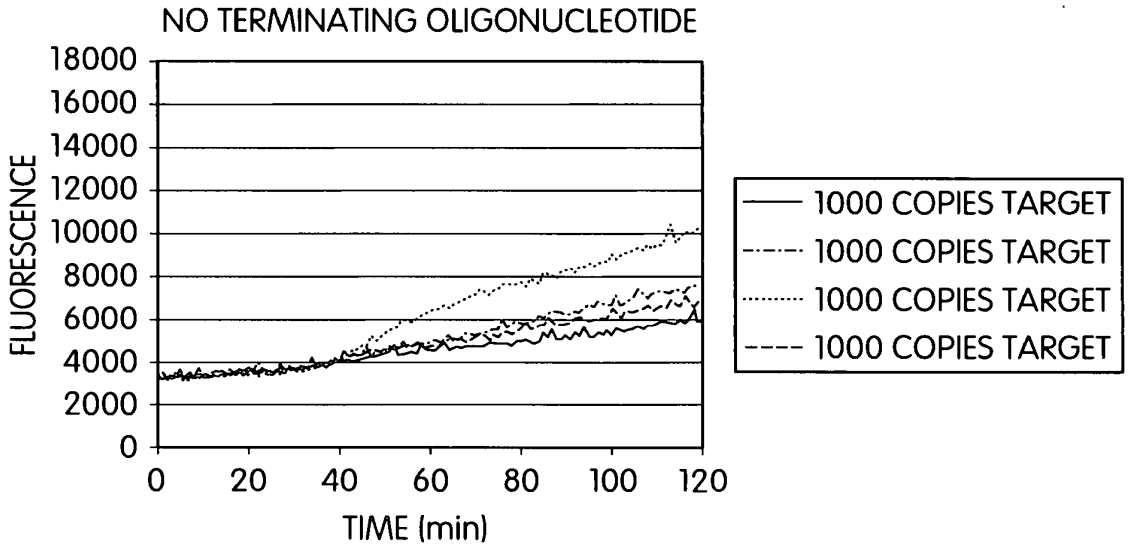

NO TERMINATING OLIGONUCLEOTIDE
FLUORESCENCE
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
120
TIME (min)
1000 COPIES TARGET
1000 COPIES TARGET
1000 COPIES TARGET
1000 COPIES TARGET

FIG. 4F

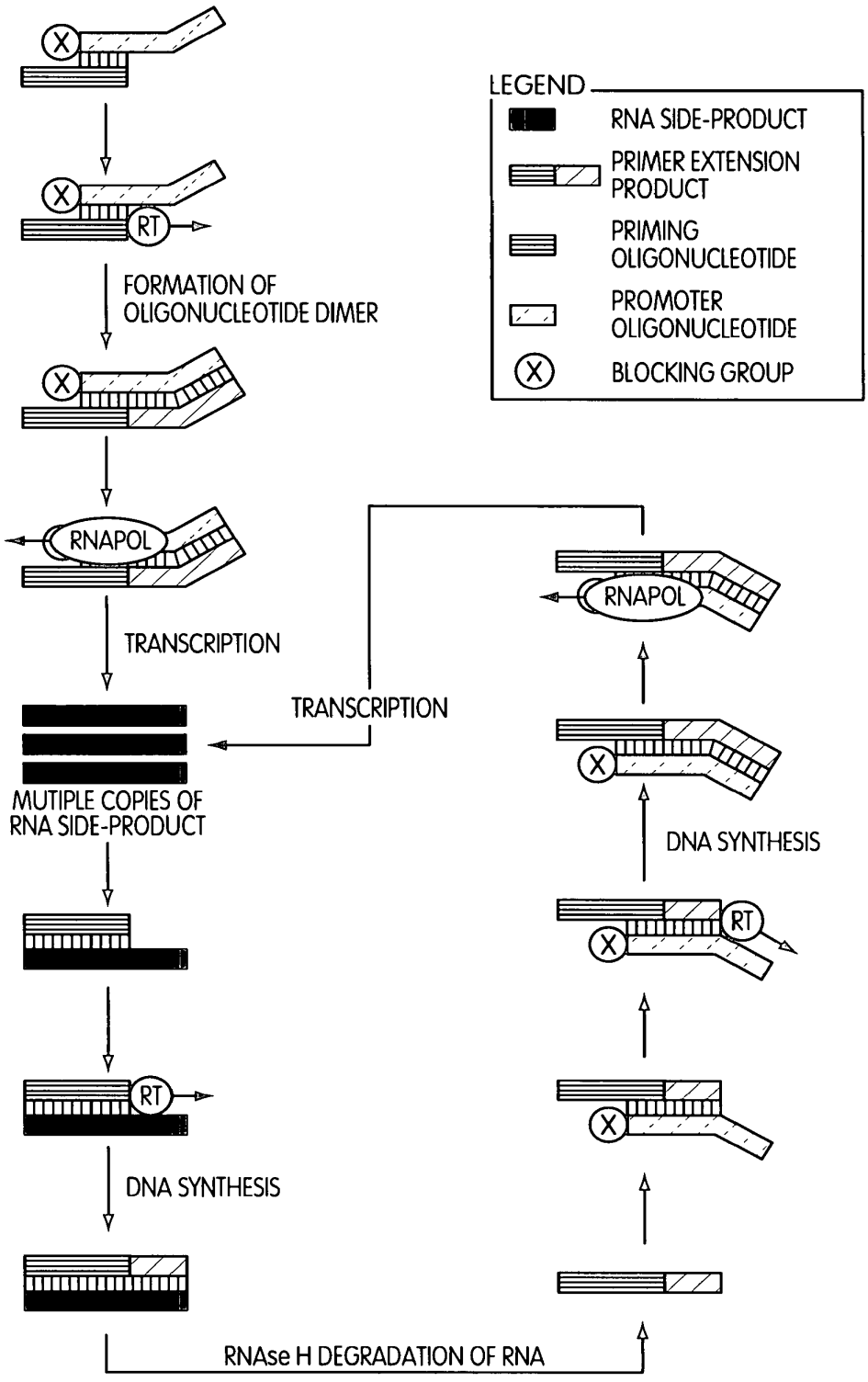


FIG. 5

SUBSTITUTE SHEET (RULE 26)

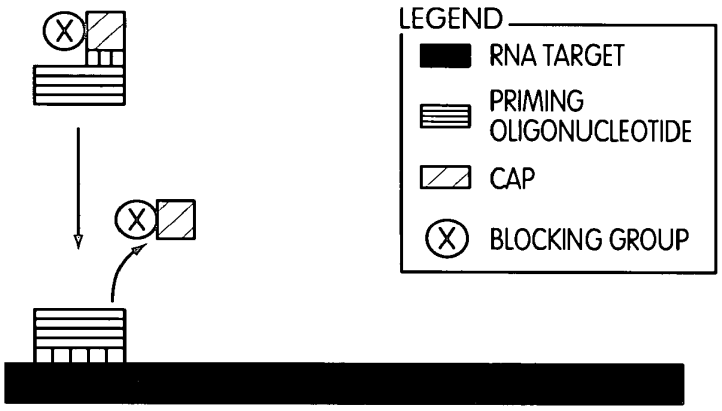

LEGEND
RNA TARGET
PRIMING OLIGONUCLEOTIDE
CAP
BLOCKING GROUP

FIG. 6A

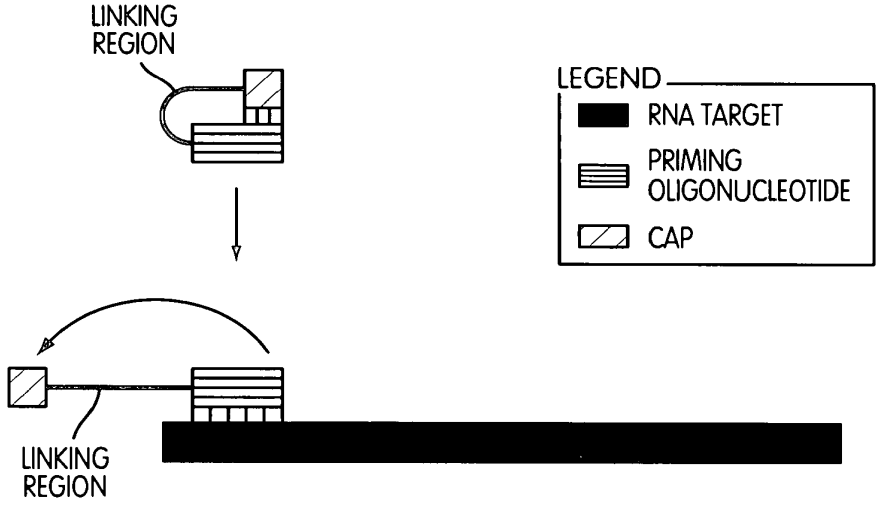

LINKING REGION
LEGEND
RNA TARGET
PRIMING OLIGONUCLEOTIDE
CAP
LINKING REGION

FIG. 6B

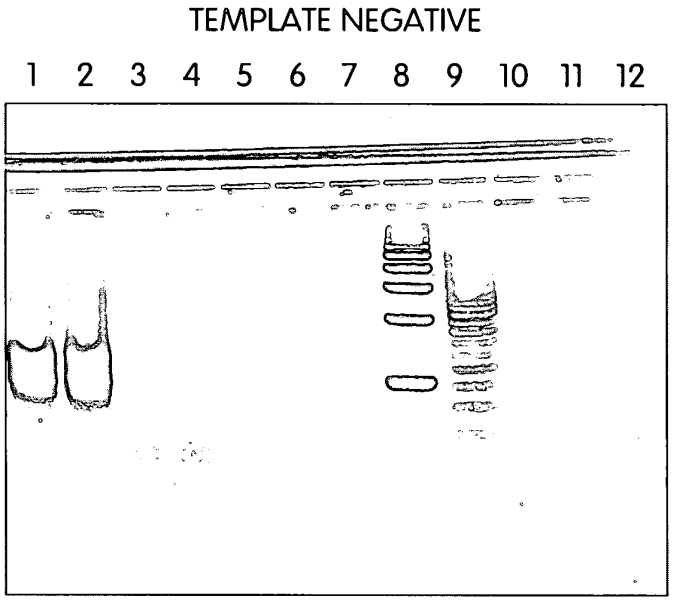

TEMPLATE NEGATIVE
1 2 3 4 5 6 7 8 9 10 11 12

FIG. 7A

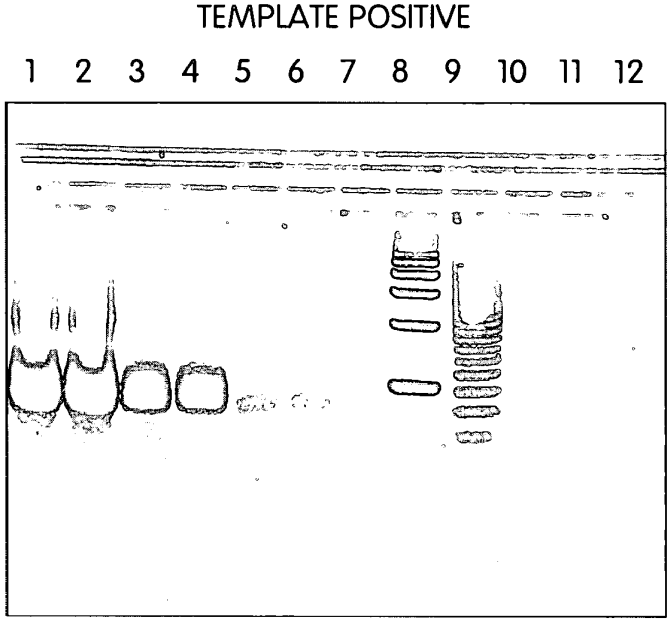

TEMPLATE POSITIVE
1 2 3 4 5 6 7 8 9 10 11 12

FIG. 7B

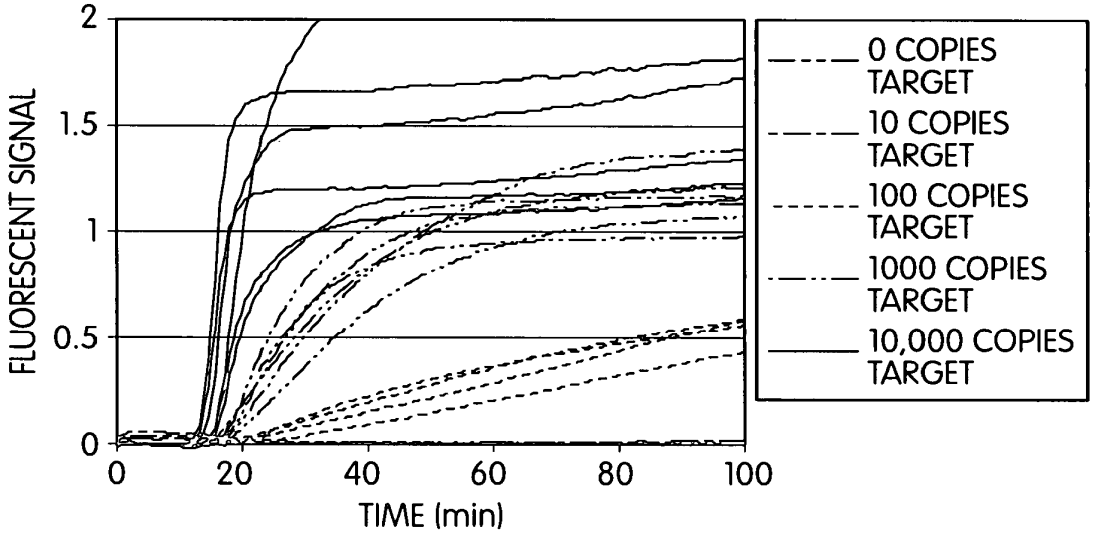

FLUORESCENT SIGNAL
2
1.5
1
0.5
0
0
20
40
60
80
100
TIME (min)
0 COPIES TARGET
10 COPIES TARGET
100 COPIES TARGET
1000 COPIES TARGET
10,000 COPIES TARGET

FIG. 8

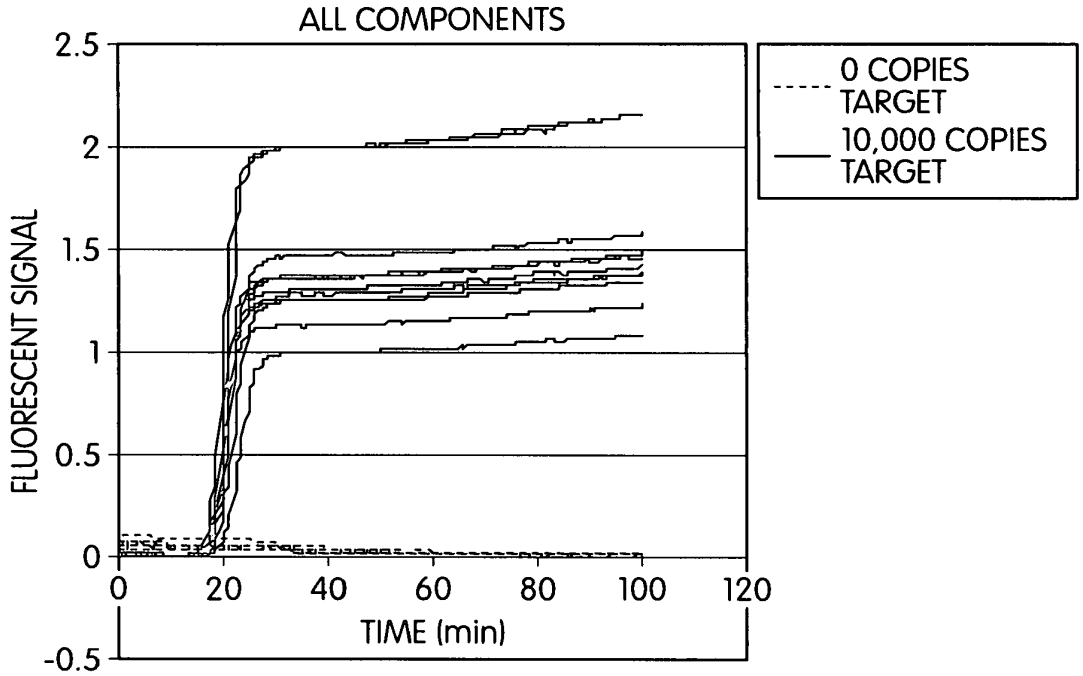

ALL COMPONENTS
FLUORESCENT SIGNAL
2.5
2
1.5
1
0.5
0
-0.5
0
20
40
60
80
100
120
TIME (min)
0 COPIES TARGET
10,000 COPIES TARGET

FIG. 9A

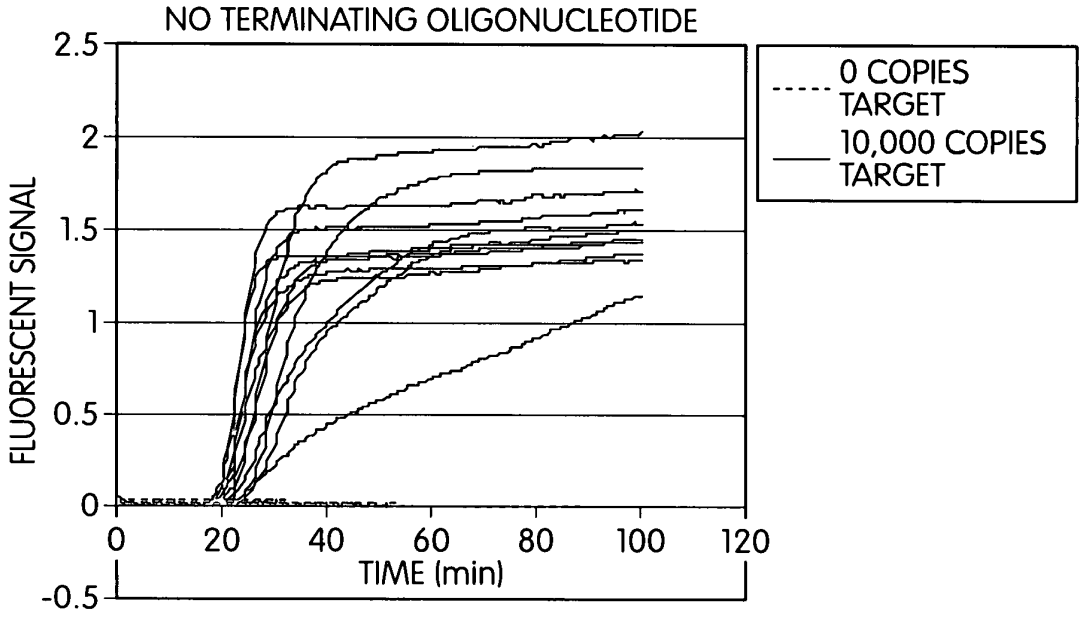

NO TERMINATING OLIGONUCLEOTIDE
FLUORESCENT SIGNAL
TIME (min)
0 COPIES TARGET
10,000 COPIES TARGET
2.5
2
1.5
1
0.5
0
-0.5
0
20
40
60
80
100
120

FIG. 9B

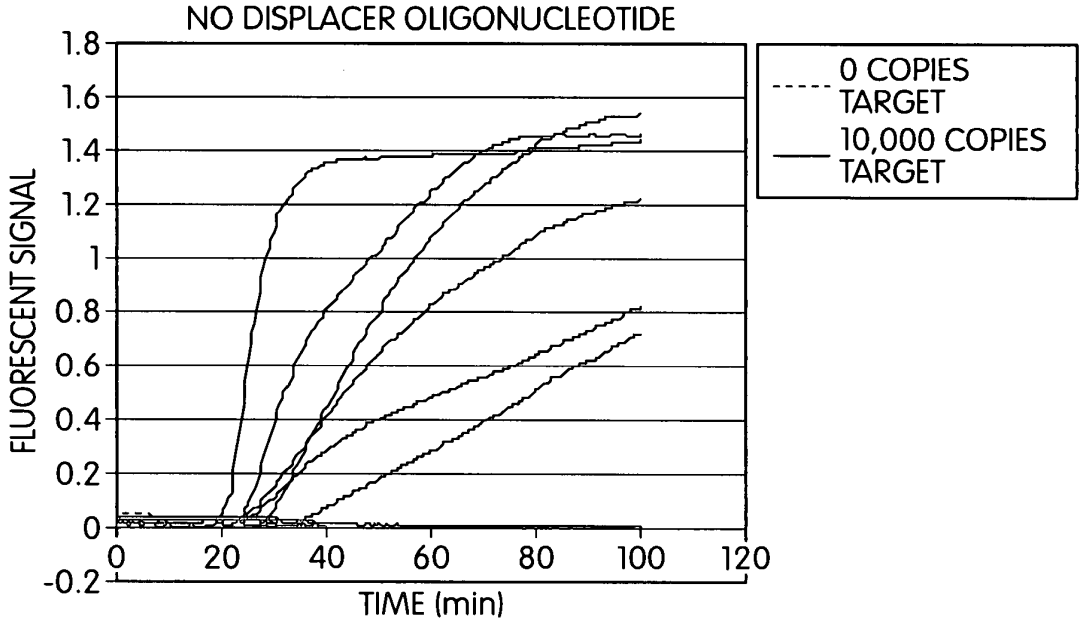

NO DISPLACER OLIGONUCLEOTIDE
FLUORESCENT SIGNAL
TIME (min)
0 COPIES TARGET
10,000 COPIES TARGET
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
20
40
60
80
100
120

FIG. 9C

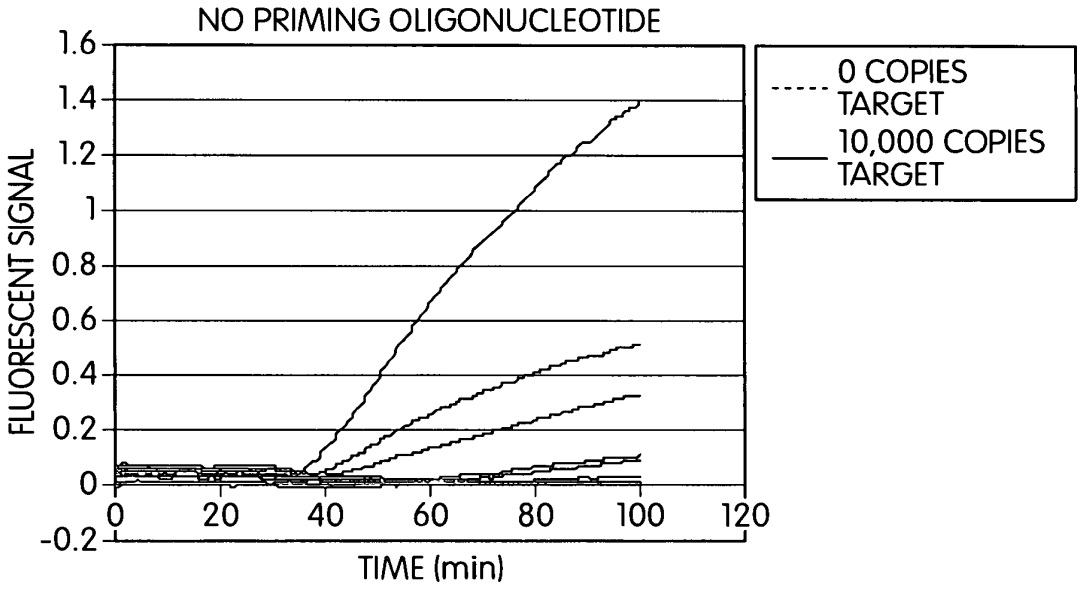

NO PRIMING OLIGONUCLEOTIDE
FLUORESCENT SIGNAL
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
20
40
60
80
100
120
TIME (min)
0 COPIES TARGET
10,000 COPIES TARGET

FIG. 9D

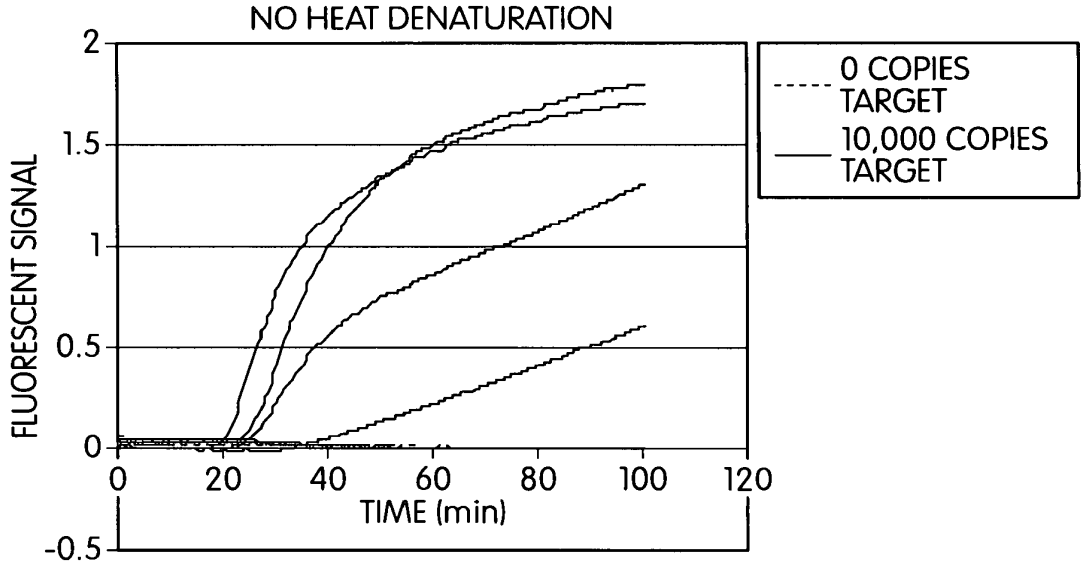

NO HEAT DENATURATION
FLUORESCENT SIGNAL
2
1.5
1
0.5
0
-0.5
0
20
40
60
80
100
120
TIME (min)
0 COPIES TARGET
10,000 COPIES TARGET

FIG. 9E

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 June 2013 (20.06.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/090360 A2

(51) International Patent Classification:
G02B 21/16 (2006.01) *G02B 27/58* (2006.01)
G02B 21/36 (2006.01)

(21) International Application Number:
PCT/US2012/069138

(22) International Filing Date:
12 December 2012 (12.12.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/576,089 15 December 2011 (15.12.2011) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **ZHUANG, Xiaowei**; 5 Luongo Farm Lane, Lexington, MA 02421 (US). **XU, Ke**; 3 Beckwith Circle, Somerville, MA 02143 (US). **BABCOK, Hazen, P.**; 5 Luongo Farm Lane, Lexington, MA 02421 (US).

(74) Agent: **OYER, Timothy, J.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: HIGH RESOLUTION DUAL-OBJECTIVE MICROSCOPY

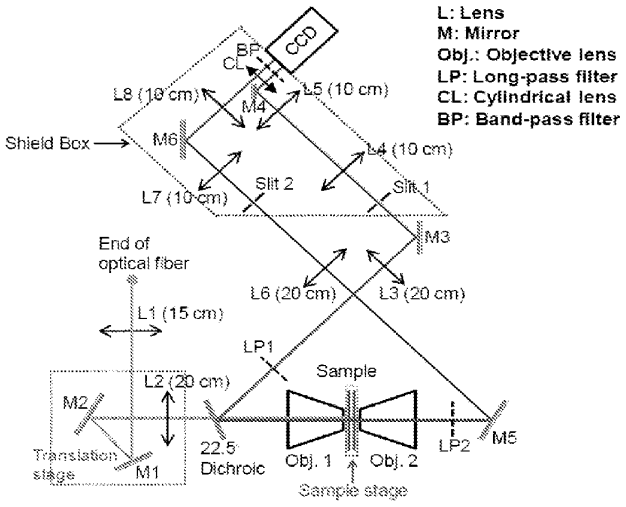


Fig. 1B

(57) **Abstract**: The present invention generally relates to super-resolution microscopy. For example, certain aspects of the invention are generally directed to a microscopy system comprising at least two objectives. In some embodiments, the microscopy system may also contain a non-circularly- symmetric lens. One or more images can be obtained using the objectives, for example, using stochastic imaging techniques such as STORM (stochastic optical reconstruction microscopy), optionally in conjunction with entities that are photoactivatable and/or photo switchable. The images obtained using the objectives may be compared, e.g., to remove noise, and/or to compare an entity present in both images, for instance, to determine the z-position of the entity. In some cases, surprisingly high resolutions may be obtained using such techniques, for example, resolutions of better than about 10 nm.

WO 2013/090360 A2

HIGH RESOLUTION DUAL-OBJECTIVE MICROSCOPY

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 61/576,089, filed December 15, 2011, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, incorporated herein by reference in its entirety.

GOVERNMENT FUNDING

Research leading to various aspects of the present invention was sponsored, at least in part, by the National Institutes of Health under Grant Nos. GM068518 and GM086214. The U.S. Government has certain rights in the invention.

FIELD OF INVENTION

The present invention generally relates to microscopy and, in particular, to super-resolution microscopy.

BACKGROUND

Super-resolution microscopy, in general, is defined as optical microscopy at resolutions that exceed the resolution limit set by the diffraction limit of light. Recent advances in super-resolution microscopy include stochastic optical reconstruction microscopy (STORM), near-field scanning optical microscopy (NSOM), stimulated emission depletion (STED), ground state depletion microscopy (GSD), reversible saturable optical linear fluorescence transition (RESOLFT), saturated structured-illumination microscopy (SSIM), and photoactivated localization microscopy (PALM). However, these techniques have certain limits on resolution, and consequently, structures below a certain size, such as certain types of biological structures, cannot be directly imaged using these optical microscopy techniques. Consequently, new optical microscopy techniques with improved resolutions are desirable.

SUMMARY

The present invention generally relates to super-resolution microscopy. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a microscopy system. In one set of embodiments, the microscopy system comprises a sample region, a first

objective on a first side of the sample region, a second objective on a second side of the sample region, and a non-circularly-symmetric lens positioned in a first imaging path in optical communication with the first objective.

In another set of embodiments, the microscopy system comprises a substantially vertically-positioned sample region, a first objective on a first side of the sample region, and a second objective on a second side of the sample region.

The microscopy system, in accordance with still another set of embodiments, includes a sample region, a first objective on a first side of the sample region, a second objective on a second side of the sample region, and means for acquiring a super-resolution image of a sample in the sample region.

The present invention is generally directed to a method, in another aspect. According to one set of embodiments, the method includes acts of acquiring a first plurality of images from a first side of a sample, acquiring a second plurality of images from a second side of the sample, and comparing the first and second plurality of images to determine positions of one or more entities in the sample by determining the shapes and/or intensities of the appearance of the entities present in the first and second plurality of images

In another set of embodiments, the method includes acts of acquiring a first plurality of images from a first side of a sample, acquiring a second plurality of images from a second side of the sample, and comparing the first and second plurality of images to determine positions of one or more entities in the sample to a resolution of less than about 1000 nm without using interference between light that forms the first plurality of images and light that forms the second plurality of images.

The method, in still another set of embodiments, includes acts of acquiring a first plurality of images from a first side of a sample, acquiring a second plurality of images from a second side of the sample, and comparing the first and second plurality of images to determine positions of one or more emissive entities in the sample, to a resolution of less than the wavelengths of the light emitted by the emissive entities, without using interference between light that forms the first plurality of images and light that forms the second plurality of images.

According to yet another set of embodiments, the method includes acts of acquiring a first plurality of images from a first side of a sample by imaging the sample

through a non-circularly-symmetric lens, and acquiring a second plurality of images from a second side of the sample.

In one set of embodiments, the method includes acts of providing a sample comprising photoswitchable fluorescent entities, activating a subset of the photoswitchable fluorescent entities, acquiring a first plurality of images from a first side of the sample, and acquiring a second plurality of images from a second side of the sample.

In another set of embodiments, the method includes acts of providing a sample comprising photoswitchable fluorescent entities, acquiring a first plurality of images from a first side of the sample using a stochastic imaging technique, and acquiring a second plurality of images from a second side of the sample using the stochastic imaging technique.

The method, in still another set of embodiments, includes acts of providing a sample comprising photoswitchable fluorescent entities, acquiring a first plurality of images from a first side of the sample, acquiring a second plurality of images from a second side of the sample, and determining x, y, and z positions of at least one of the photoswitchable fluorescent entities in the sample, using the first and second plurality of images, to a resolution of less than about 1000 nm. In another set of embodiments, the method includes acts of providing a sample comprising photoswitchable fluorescent entities, acquiring a first plurality of images from a first side of the sample, acquiring a second plurality of images from a second side of the sample, and determining x, y, and z positions of at least one of the photoswitchable fluorescent entities in the sample, using the first and second plurality of images, to a resolution of less than a wavelength of light emitted by the photoswitchable fluorescent entity.

According to yet another set of embodiments, the method includes acts of providing a sample comprising one or more entities, acquiring a first plurality of images from a first side of the sample, acquiring a second plurality of images from a second side of the sample, accepting entities due to anticorrelated changes between the appearance of the entities in the first plurality of images and the appearance of the entities in the second plurality of images, and assembling the accepted entities into a final data set or image. The method, in still another set of embodiments, includes acts of providing a sample comprising one or more entities, acquiring a first plurality of images from a first side of

the sample, acquiring a second plurality of images from a second side of the sample, rejecting entities due to correlated changes between the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images, and assembling the first and second plurality of images into a final data set or image while suppressing the rejected entities.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1B illustrate various configurations of microscopy systems in accordance with certain embodiments of the invention;

Figs. 1C-1E illustrate super-resolution imaging, in certain embodiments of the invention;

Figs. 2A-2F illustrate super-resolution imaging of individual actin filaments in cells, in certain embodiments of the invention; and

Figs. 3A-3M illustrate super-resolution imaging of actin networks in cells, in accordance with some embodiments of the invention.

DETAILED DESCRIPTION

The present invention generally relates to super-resolution microscopy. For example, certain aspects of the invention are generally directed to a microscopy system comprising at least two objectives. In some embodiments, the microscopy system may also contain a non-circularly-symmetric lens. One or more images can be obtained using the objectives, for example, using stochastic imaging techniques such as STORM (stochastic optical reconstruction microscopy), optionally in conjunction with entities that are photoactivatable and/or photoswitchable. The images obtained using the objectives may be compared, e.g., to remove noise, and/or to compare an entity present in both images, for instance, to determine the z-position of the entity. In some cases, surprisingly high resolutions may be obtained using such techniques, for example, resolutions of better (or less) than about 10 nm in terms of full width at half maximum.

In some aspects, the present invention is generally directed to microscopy systems, especially optical microscopy systems, for acquiring images at super-resolutions, or resolutions that are smaller than the theoretical Abbe diffraction limit of light. In certain embodiments of the invention, as discussed below, surprisingly high (small) resolutions may be obtained using such techniques, for example, resolutions of less than about 20 nm, less than about 15 nm, or less than about 10 nm.

One example of an embodiment of the invention is now described with respect to Fig. 1. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. Fig. 1A illustrates microscopy system 10 comprising sample region 20 containing sample 25, illumination source 30, and detector 50. On either side of sample 25 are objectives 41 and 42. Light 35 from illumination source 30 may be directed to sample region 20 to illuminate sample 25. For instance, light from the illumination source may be directed through one or more of the objectives, or as is shown in Fig. 1A, light 35 may be used to illuminate sample region 20 without being passed through objectives 41 or 42. In some cases, the light from the illumination source may interact with various optical components, such as lenses, mirrors (e.g., dichroic mirrors or polychroic mirrors), beam splitters, filters, slits, windows, prisms, diffraction gratings, optical fibers, etc., before illuminating the sample. In some cases, more than one illumination source can be used. In some cases, the light from the sample being collected on the first side using the first objective and the light from the sample

being collected on the second side using the second objective may be directed onto two separate detectors, instead of the single detector 50 as shown in Fig. 1A.

One or more images may be acquired using one or more detectors and/or one or more objectives. For example, as is shown in Fig. 1A, imaging paths 45 and 46 from sample 25 respectively pass through objectives 41 and 42 before reaching detector 50. In this figure, mirrors 47, 48 are used to direct each of the imaging paths to detector 50. However, in other examples, a variety of optical components may be used to direct the imaging paths to a detector (or to different detectors), for example, lenses, mirrors, beam splitters, filters, slits, windows, prisms, diffraction gratings, optical fibers, etc.

In one set of embodiments, super-resolution techniques may be used to obtain one or more images from sample 25, via one or both of objectives 41 and 42. For example, a stochastic imaging technique such as STORM ("Stochastic Optical Reconstruction Microscopy") or 3D-STORM may be used. See, e.g., International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*, published as WO 2009/085218 on July 9, 2009; and U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, each incorporated herein by reference. In some stochastic imaging techniques, incident light is applied to a sample to cause a statistical subset of entities present within the sample to emit light, the emitted light is acquired or imaged, and the entities are deactivated (either spontaneously, or by causing the deactivation, for instance, with suitable deactivation light). This process may be repeated any number of times, each time causing a statistically different subset of the entities to emit light, and this process may be repeated to produce a final, stochastically produced image.

If two or more objectives are used, as is shown in Fig. 1A, then one or more images may be obtained with each of the objectives. In some cases, the images obtained using the objectives are compared to determine correlations between the images. For example, the same emissive entity may be observed in a first image obtained using a first objective and a second image using a second objective. In certain embodiments, the shape and/or intensity of the entity can be compared, for example, to determine the location of the entity, and/or to determine whether to accept or reject the entity.

In one set of embodiments, the shape and/or intensity of the entity in the first and second images are compared to determine the position of the entity within the sample region, for example, if a non-circularly-symmetric lens is used. In some embodiments, using certain types of non-circularly-symmetric lenses, the shape and/or intensity of an entity (i.e., of the appearance of the entity in an image) may be distorted to different degrees based in part on the distance between the entity and the focal plane of the objective (i.e., in the z direction). Accordingly, by observing the amount of distortion present in the first and/or second images, the position of the entity within the sample region may be determined. Using two objectives, in some cases, can result in significantly higher resolution of the entity in the z direction, than using just a single objective.

In certain embodiments, a property of the entity in a first image detected using the first objective and in a second image detected using the second objective can be compared to determine if that property appears to be correlated or anticorrelated, and the entity may be accepted or rejected based on the anticorrelated or correlated appearance of the entity in the images. For example, the property may be the shape and/or intensity of the entity in the first image and in the second image. Thus, for instance, an anticorrelated entity can appear to be distorted in a first direction in the first image and distorted in a second direction in the second image, where the second direction is in a different direction than the first direction; an entity that is not anticorrelated may be an entity in the first and second images that has distortions in each image in substantially the same direction, or the entity may appear to be undistorted in one or both images, etc. Thus, for example, anticorrelated entities may be accepted as "true" entities present within the sample region, while entities that do not show such an anticorrelated appearance may be rejected as being noise or abnormalities.

While others have used super-resolution microscopy and dual-objective imaging to determine the positions of entities in the z direction in a sample region, those techniques typically rely on interferometry rather than comparing the image shape and/or intensity of the entities detected from the two objectives to obtain information about the position of the entity in the z direction. Such techniques are cumbersome and require optical components and configurations necessary to cause a photon to form its own coherent reference beam in order to interfere with itself to produce a signal that can be

analyzed via interferometry, such that information about the z direction can be obtained. For example, in order to cause a photon to form its own coherent reference beam in order to interfere with itself, one or more beam splitters must be used, and/or the optics must be sufficiently precisely aligned so that light from the sample region obtained from the two objectives interferes or otherwise superposes with itself. Accordingly, very precise alignments are needed. However, as discussed herein, in various embodiments of the present invention, interference between light collected using various objectives is unnecessary to determine the positions of entities in the z direction; instead, information about the z direction is determined directly from the images of the entities obtained from the two objectives, for example, by comparing shape and/or intensity information of the images detected using the two objectives.

As mentioned, certain aspects of the present invention are directed to microscopy systems, especially optical microscopy systems, able to produce super-resolution images (or data sets). In some cases, the microscopy system comprises a plurality of objectives, e.g., positioned on various sides of the sample region, and a non-circularly-symmetric lens positioned in at least a first imaging path in optical communication with at least one of the objectives. Such microscopy systems may be used to study any of a variety of suitable samples. The samples can be biological and/or non-biological in origin. For example, the sample studied may be a non-biological sample (or a portion thereof) such as a microchip, a MEMS device, a nanostructured material, or the sample may be a biological sample such as a cell, a tissue, a virus, or the like (or a portion thereof).

The microscopy system may comprise a sample region for holding and/or containing a sample. In some cases, the sample region is substantially planar, although in other cases, a sample region may have other shapes. In certain embodiments, the sample region (or the sample contained therein) has an average thickness of less than about 1 mm, less than about 300 micrometers, less than about 100 micrometers, less than about 30 micrometers, less than about 10 micrometers, less than about 3 micrometers, less than about 1 micrometer, less than about 750 nm, less than about 500 nm, less than about 300 nm, or less than about 150 nm.

In some cases, as is discussed below, the sample region is substantially vertically positioned. Typically, one or more objectives can be positioned relative to the sample region such that a focal plane of an objective is also substantially vertically positioned.

When a sample is positioned vertically in the sample region, the force of gravity on the sample can cause entities in the sample to move or settle in the same focal plane. Accordingly, entities moving under the effects of gravity will continue to stay in focus in images obtained using the objective, unlike more traditionally horizontally-positioned sample regions, where even slight changes in the positions of the entities under the effects of gravity could potentially cause the entities to move into or out of the focal plane of the objective, accordingly making it harder to determine the position of the entities. In certain cases, for instance, the correction of movement or drift of entities within the focal plane of the objective (i.e., in the *x*-*y* directions) is more precise than in the *z* direction substantially normal to the focal plane of the objective. Accordingly, a vertically-positioned sample region allows entities to be more precisely studied, as movement of the entities due to gravity does not cause the entities to move out of the focal plane of the objective. However, it should be understood that vertical positioning of a sample region is not a requirement of the present invention, and in other embodiments, the sample region may be substantially horizontally positioned instead.

In certain embodiments, immersion objectives are used, for instance, oil immersion lenses, water immersion lenses, solid immersion lenses, etc. (although in other embodiments, other, non-immersion objectives can be used). In certain embodiments, objectives are positioned on various sides of a sample. If immersion objectives are used, it may be more difficult to position the sample such that forces exerted by the immersion fluid on the sample (e.g., due to surface tension, capillary action, etc.) are able to substantially balance each other in a horizontally-positioned sample (for example, due to differing gravitational effects on the movement of fluids and/or the sample). However, if the sample is positioned substantially vertically, the forces created by the immersion fluids on either side of the sample may be substantially equal (since gravity would not play a major role in these forces), e.g., if the fluid compositions and/or amounts are the same. Accordingly, there would be less of a tendency for the sample to move in a particular direction within the sample region, thereby improving precision or resolution of the images of the sample.

Any of a variety of techniques can be used to position a sample within the sample region (which may be substantially horizontally positioned, substantially vertically positioned, or positioned at any other suitable angle). For example, the sample may be

positioned in the sample region using clips, clamps, or the like. In some cases, the sample can be held or manipulated using various actuators or controllers, such as piezoelectric actuators. Suitable actuators having nanometer precision can be readily obtained commercially. For example, in certain embodiments, the sample may be positioned relative to a translation stage able to manipulate at least a portion of the sample region, and the translation stage may be controlled at nanometer precision, e.g., using piezoelectric control.

As previously discussed, two or more objectives may be used, in accordance with various embodiments of the invention. The objectives may each be any suitable objective, and may each be air or immersion objectives. The objectives may each independently be the same or different. The objectives can have any suitable magnification and any suitable numerical aperture, although higher magnification objectives are typically preferred. For example, the objectives may each be about 4x, about 10x, about 20x, about 32x, about 50x, about 64x, about 100x, about 120x, etc., while in some cases, the objective may have a magnification of at least about 50x, at least about 80x, or at least about 100x, The numerical aperture can be, for instance, about 0.2, about 0.4, about 0.6, about 0.8, about 1.0, about 1.2, about 1.4, etc. In certain embodiments, the numerical aperture is at least 1.0, at least 1.2, or at least 1.4. Many types of microscope objectives are widely commercially available.

Any number of objectives may be used in different embodiments of the invention, and the objectives may each independently be the same or different, depending on the application. In one set of embodiments, for instance, two objectives are used, positioned on either side of a sample region. The objectives can be positioned such that each objective is used to image the same location of the sample region (or at least, such that the regions imaged by each of the objectives overlaps). Such a configuration is also referred to as a "dual-objective" system. For instance, the objectives may be positioned at about 180° relative to each other, or at any other suitable angles such that each objective can focus on the same location within the sample region. The objectives can also be positioned in some embodiments such that focal planes for each objective overlap, and/or such that the objectives are collinearly positioned relative to each other.

However, in other embodiments, other numbers of objectives may be used. For example, a microscopy system may have three, four, five, six, etc. objectives that can be used to image a sample in the sample region. For example, a microscopy system may have at least three, four, five, six, etc. objectives each focused on a single position within the sample region (e.g., such that the regions imaged by each of the objectives overlaps), or positioned such that there are multiple focal positions within the sample region for various objectives. As a non-limiting example, a microscopy system may have four objectives and two focal positions, each of which is focused on by two objectives.

As will be discussed in more detail below, in certain embodiments, microscopy systems such as those discussed herein may be used for locating the z position of entities within a sample region. The z position is typically defined to be in a direction defined by an objective (e.g., towards or away from the objective). In some cases, the z position can also be orthogonal to the focal (x-y) plane of the objective. The sample is usually substantially positioned within the focal plane of the objective, and thus, the z direction may also be taken in some embodiments to be in a direction substantially normal to the sample or the sample region (or at least a plane defined by the sample, e.g., if the sample itself is not substantially flat), e.g., in embodiments where the sample and/or the sample region is substantially planar.

In one set of embodiments, the z position of an entity within a sample region is determined in microscopy systems having a non-circularly-symmetric lens, using techniques such as astigmatism imaging or any other suitable imaging technique. The z position of an entity within a sample region may also be determined in microscopy systems with or without a non-circularly-symmetric lens using techniques such as off-focus imaging, multi-focal plane imaging, or any other suitable imaging technique. Typically, a non-circularly-symmetric lens is a lens that is not circularly symmetric with respect to the direction light emitted from a sample passes through the lens. For instance, the lens may be cylindrical, ellipsoidal, or the like. In some cases, the lens may have different radii of curvature in different planes. The cylindrical lens may also be a weak cylindrical lens, e.g., having a relatively long focal length, in certain embodiments. For example, the cylindrical lens may have a focal length of 100 mm or 1 m. In some embodiments, the non-circularly-symmetric lens can also be positioned relative to a sample region to define a focal region where at least a portion of the focal region does

not contain the sample region. Light from an entity in a sample passing through a non-circularly symmetric lens may appear in an acquired image to be circular or elliptical. Non-circularly-symmetric lens may be obtained from various commercial sources.

The non-circularly-symmetric lens may be positioned on an optical or an imaging path extending from the sample region through an objective to a suitable detector, as discussed below. In some cases, more than one imaging path can pass through a non-circularly-symmetric lens. As an example, as is shown in Fig. 1B, two imaging paths, each of which extend through the sample region and one of the objectives positioned on either side of the sample region, are passed through a common non-circularly-symmetric lens ("CL") before reaching a common detector ("CCD").

In still other embodiments, there may be more than one non-circularly-symmetric lens present. As an example, if more than one detector is used, there may be non-circularly-symmetric lenses in imaging paths to some or all of the detectors. If more than one non-circularly-symmetric lens is present, the lenses may each independently be the same or different. In some cases, a non-circularly-symmetric lens can be positioned in a first imaging path in optical communication with a first objective, and optionally, a non-circularly-symmetric lens can also be positioned in a second imaging path in optical communication with a second objective.

The imaging path is not necessarily a straight line, although it can be in certain instances. The imaging path can be any path leading from the sample region, optionally through one or more optical components, to a detector such that the detector can be used to acquire an image of the sample region. Any of a variety of optical components may be present, and may serve various functions. For example, optical components may be present to guide the imaging path around the microscopy system, to reduce noise or unwanted wavelengths of light, or the like. Non-limiting examples of optical components that may be present within the imaging path (or elsewhere in the microscopy system, such as in an illumination path between a source of illumination and a sample region) include one or more optical components such as lenses, mirrors (for example, dichroic mirrors, polychroic mirrors, one-way mirrors, etc.), beam splitters, filters, slits, windows, prisms, diffraction gratings, optical fibers, and any number or combination of these may be present in various embodiments of the invention. One non-limiting example of a microscopy system containing several optical components in various

imaging paths between a sample region through various objectives to a common detector is shown in Fig. 1B, and is discussed in more detail below.

The detector can be any device able to acquire one or more images of the sample region, e.g., via an imaging path. For example, the detector may be a camera such as a CCD camera, a photodiode, a photodiode array, a photomultiplier, a photomultiplier array, a spectrometer, or the like. The detector may be able to acquire monochromatic and/or polychromatic images, depending on the application. Those of ordinary skill in the art will be aware of detectors suitable for microscopy systems, and many such detectors are commercially available.

In one set of embodiments, a single detector is used, and multiple imaging paths may be routed to the common detector using various optical components such as those described herein. A common detector may be advantageous, for example, since no calibration or correction may need to be performed between multiple detectors. For instance, with a common detector, there may be no need to correct for differences in intensity, brightness, contrast, gain, saturation, color, etc. between different detectors. Thus, for example, a first image of the sample region can be projected onto a first location of the detector via a first imaging path, while a second image of the sample region can be projected onto a second location of the detector via a second imaging path. In some embodiments, images may be acquired by the detector simultaneously, e.g., as portions of the same overall frame acquired by the detector. This could be useful, for instance, to ensure that the images from the various objectives are properly synchronized with respect to time. Note that while these images may be acquired on the same frame within the detector, the images from the separate objectives are nevertheless generally referred to herein as separate images, although the separate images may represent different views of a common sample, or of a region within the common sample.

However, in other embodiments of the invention, more than one detector may be used, and the detectors may each independently be the same or different. In some cases, multiple detectors may be used, for example, to improve resolution and/or to reduce noise. For example, at least 2, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, etc. detectors may be used, depending on the application. For example, a microscopy system can comprise a first detector in optical communication with a first objective via a first imaging path and a second detector in optical

communication with a second objective via a second imaging path. This may be useful, for example, to simplify the collection of images via different imaging paths from different sides of the sample region. In some cases, more than two detectors may be present within the microscopy system.

The sample region is illuminated, in certain embodiments of the invention, using an illumination source that is able to illuminate at least a portion of the sample region via one or more illumination paths. Like the imaging path, the illumination path need not be a straight line, but may be any suitable path leading from the illumination source, optionally through one or more optical components, to at least a portion of the sample region.

In some embodiments, a portion of the illumination path may also coincide with a portion of the imaging path. As a non-limiting example, as is illustrated in Fig. 1B, part of an illumination path between an illumination source (at the end of the optical fiber) and the sample region passes through Objective 1 ("Obj. 1") before reaching the sample region, while one of the imaging paths likewise passes through Objective 1. The imaging path and the illumination path proceed in different directions at a dichroic mirror. However, this is not a requirement, as the example in Fig. 1A shows. Accordingly, at least some of the optical components that the illumination path passes through may be the same or different than the imaging paths. Non-limiting examples of optical components include any number of lenses, mirrors, beam splitters, filters, slits, windows, prisms, diffraction gratings, optical fibers, etc.

The illumination source may be any suitable source able to illuminate at least a portion of the sample region. The illumination source can be, e.g., substantially monochromatic or polychromatic. The illumination source may also be, in some embodiments, steady-state or pulsed. In some cases, the illumination source produces coherent light. In one set of embodiments, at least a portion of the sample region is illuminated with substantially monochromatic light, e.g., produced by a laser or other monochromatic light source, and/or by using one or more filters to remove undesired wavelengths. In some cases, more than one illumination source may be used, and each of the illumination sources may be the same or different. For example, in some embodiments, a first illumination source may be used to activate entities in a sample region, and a second illumination source may be used to excite entities in the sample

region, or to deactivate entities in the sample region, or to activate different entities in the sample region, etc.

In one set of embodiments, the illumination path may reach the sample region at incidence angles at or slightly smaller than the critical angle for total internal reflection at the interface between the sample region and a glass substrate, e.g., a glass coverslide. In some cases, the illumination path can reach the sample region at incidence angles at or slightly smaller than the critical angle of the glass-water interface between the objective and the sample region. For instance, the incidence angle may be less than about 90°, less than about 80°, less than about 70°, less than about 60°, less than about 50°, less than about 40°, less than about 30°, less than about 20°, or less than about 10°. Typically, the incidence angle is defined as the angle between the light propagation direction and the direction normal to the sample.

In certain aspects, two or more objectives are used to obtain one or more images of a sample in a sample region. For example, a first plurality of images of a first side of a sample may be obtained via one objective, and a second plurality of images of a second side of the sample may be obtained via a second objective. In some cases, as previously discussed, more than two objectives may be present, e.g., focused on the same or different regions of the sample. In certain embodiments, an image from the first plurality of images and an image from the second plurality of images are compared to determine a position of an entity in the sample, e.g., if these two were obtained simultaneously or substantially simultaneously. Various properties of an entity within the first and second plurality of images can be compared. For instance, properties such as the position, shape, size, color, intensity, parallax, and/or appearance of the entity in the images can be compared. As specific non-limiting examples, the ellipticity of the appearance of an entity may be determined, or the degree of focus of the entity in the image may be determined.

For example, in one set of embodiments, an entity visible in both of the first and second pluralities of images, each obtained from different sides or angles of the sample, may be compared to determine the appearance of the entity in each of the pluralities of images. The images can be recorded by the same or different detectors, and in some cases, some or all of the images may be obtained simultaneously or substantially simultaneously. In some cases, based on this comparison, an entity can be accepted or

rejected. For example, if an entity appears in a first image (e.g., from a first objective) but not in a second image (e.g., from a second objective), it may be determined that the entity is an artifact, and the entity could thereby be rejected. As another example, if the entity appears in both images, and appears to be in focus in both images (for example, as a single point, or a point with low spread in the image), it may be determined that the entity is in the focal plane of both objectives, and the entity can be accepted or rejected on that basis. As yet another example, if the entity is in focus in one image but not in another image, it may be determined that the entity is an artifact or "noise," and the entity could thereby be rejected on that basis.

As another example, in some embodiments, an entity may be accepted or rejected based on a correlated or an anticorrelated property of the appearance of the entity in the first plurality of images formed by the first objective and the appearance of the entity in the second plurality of images formed by the second objective. For example, in one set of embodiments, a property of an entity in first and second plurality images is compared to determine if the property appears to be correlated or anticorrelated, and the entity can be accepted or rejected based on the anticorrelated or correlated property of the entity. Generally, an anticorrelated property is one that has a first appearance in a first image and an inversely-related appearance in a second image. For example, If the image of the entity formed by the first objective and the image of the entity formed by the second objective both appear elongated in the same direction, it may be that the property is correlated and the entity is an artifact or "noise," and the entity could thereby be rejected on that basis.

Thus, according to certain embodiments, the noise level of images may be substantially reduced, thereby improving the precision or resolution of the final image (or data set). The amount of noise in the images can affect the quality of the images obtained using such techniques, and by removing such sources of noise, the resolution may be improved. By using techniques such as those described herein to eliminate a major source of uncertainty, the resolution of the final images (or data sets) may be significantly enhanced.

As a specific non-limiting example, as discussed below, the appearance of the shape and/or intensity of an entity in an image may be related to the position of the entity in the z direction, for instance, if a non-circularly-symmetric lens such as a cylindrical

lens is used. Thus, for example, the ellipticity or elongated shape of the image of the entity may be a function of the distance between the entity and the focal plane of the objective. Accordingly, by determining the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images, the position of the entity in the *z* direction can be determined.

In some embodiments, starting with a plurality of images (e.g., a movie or a data set), some or all of the entities may be identified (e.g., through fluorescence, phosphorescence, etc.), and the positions of these entities can be determined. For example, the appearance of the entities in the images can be fit, in some cases, to Gaussian and/or elliptical Gaussian functions to determine their centroid positions, intensities, widths, ellipticities, etc. By determining the center or centroid position of the image of an entity using various techniques, for example, using average locations or least-squares fitting to a 2-dimensional Gaussian function of the intensity profile of the image, the location of the entity in the sample can be determined in the directions parallel to the focal plan (*x* and *y*), typically at a high (small) resolution. By determining the ellipticity of the image, for example, using least-squares fitting to a 2-dimensional Gaussian function of the intensity profile of the image, or to other suitable distributions, the location of the entity in the sample can be determined in the direction perpendicular to the focal plan (*z*), typically at a high (small) resolution. This process can be repeated as necessary for any or all of the entities within the image.

A final image or data set may be assembled or constructed from the positions of the entities or a subset of entities in the sample in some embodiments of the invention. In some cases, the data set may include position information of the entities in the *x*, *y*, and optionally *z* directions. As an example, the final coordinates of an entity may be determined as the average of the position of the entity as determined using each of the objectives, or as a weighted average of the position of the entity as determined using each of the objectives (e.g., weighted by the width of the image and/or number of photons obtained by each objective, etc.). The entities may also be colored in a final image in some embodiments, for example, to represent the degree of uncertainty, to represent the location of the entity in the *z* direction, to represent changes in time, etc. In one set of embodiments, a final image or data set may be assembled or constructed based

on only the locations of the accepted entities while suppressing or eliminating the locations of the rejected entities.

As previously discussed, z direction information about entities within a sample may be obtained in certain embodiments. In some cases, the z positions can be determined at a resolution that is less than the diffraction limit of the incident light. In one set of embodiments, the emitted light may be processed, using Gaussian fitting, linear averaging, or other suitable techniques to localize the position of the emissive entities, e.g., as discussed herein. For example, for visible light, the z position of an entity can be determined at a resolution less than about 1000 nm, less than about 800 nm, less than about 500 nm, less than about 300, less than about 200 nm, less than about 100 nm, less than about 50 nm, less than about 40 nm, less than about 35 nm, less than about 30 nm, less than about 25 nm, less than about 20 nm, less than about 15 nm, or less than about 10 nm, using techniques such as these.

Any microscopy technique able to determine the z position of entity in a sample may be used in various embodiments of the invention. Non-limiting examples include astigmatism imaging, off-focus imaging, multi-focal plane imaging, or the like. In some cases, the entity may be positioned and imaged such that the entity does not appear as a single point of light, but as an image that has some area, for example, as a slightly unresolved or unfocused image. As a specific example, an entity can be imaged by a lens or a detector system that defines one or more focal regions (e.g., one or more focal planes) that do not contain the entity, such that the image of the entity at the detector appears unfocused. The degree to which the entity appears unfocused can be used to determine the distance between the entity and one of the focal regions, which can then be used to determine the z position of the entity.

In one set of embodiments, the z position can be determined using astigmatism imaging, for example, using a non-circularly-symmetric lens, as previously discussed. The size, shape, ellipticity, etc. of the image of an entity can be used, in some cases, to determine the distance between the entity and the focal region of the lens or the detector, which can be used to determine the z position of the entity in the sample. In some cases, the appearance of entities that are out of focus appear to be increasingly elliptical with distance from the focal plane, with the direction of ellipticity indicating whether the entity is above or below the focal plane.

In another set of embodiments, the z position can be determined using off-focus imaging. An entity not in one of the focal planes defined by an objective may appear to be unfocused, and the degree that the entity appears unfocused may be used to determine the distance between the entity and the focal plane, which can then be used to determine the z position. In some cases, the image of the unfocused entity may appear generally circular (with the area being indicative of the distance between the entity and the focal region of the lens), and in some instances, the image of the unfocused entity can appear as a series of ring-like structures, with more rings indicating greater distance).

In yet another set of embodiments, e.g., with multi-focal plane imaging, the light emitted by the entities may be collected by a plurality of detectors. In some cases, at one or more of the detectors, the light may appear to be unfocused. The degree that the images appear unfocused can be used to determine the z position in certain embodiments of the invention.

As previously discussed, imaging techniques such as these avoid interferometry, which typically requires optical components and configurations necessary to cause a photon to form its own coherent reference beam in order to interfere with itself. In contrast, optical techniques such as those described herein do not generally require precise alignment of a coherent reference beam.

In certain aspects of the invention, images of a sample may be obtained using stochastic imaging techniques. In many stochastic imaging techniques, various entities are activated and emit light at different times and imaged; typically the entities are activated in a random or "stochastic" manner. For example, a statistical or "stochastic" subset of the entities within a sample can be activated from a state not capable of emitting light at a specific wavelength to a state capable of emitting light at that wavelength. Some or all of the activated entities may be imaged (e.g., upon excitation of the activated entities), and this process repeated, each time activating another statistical or "stochastic" subset of the entities. Optionally, the entities are deactivated (for example, spontaneously, or by causing the deactivation, for instance, with suitable deactivation light). Repeating this process any suitable number of times allows an image of the sample to be built up using the statistical or "stochastic" subset of the activated emissive entities activated each time. Higher resolutions may be achieved in some cases because the emissive entities are not all simultaneously activated, making it easier to

resolve closely positioned emissive entities. Non-limiting examples of stochastic imaging which may be used include stochastic optical reconstruction microscopy (STORM), single-molecule localization microscopy (SMLM), spectral precision distance microscopy (SPDM), super-resolution optical fluctuation imaging (SOFI), photoactivated localization microscopy (PALM), and fluorescence photoactivation localization microscopy (FPALM).

In certain embodiments, the resolution of the entities in the images can be, for instance, on the order of 1 micrometer or less, as described herein. In some cases, the resolution of an entity may be determined to be less than the wavelength of the light emitted by the entity, and in some cases, less than half the wavelength of the light emitted by the entity. For example, if the emitted light is visible light, the resolution may be determined to be less than about 700 nm. In some cases, two (or more) entities can be resolved even if separated by a distance of less than about 500 nm, less than about 300 nm, less than about 200 nm, less than about 100 nm, less than about 80 nm, less than about 60 nm, less than about 50 nm, or less than about 40 nm. In some cases, two or more entities separated by a distance of less than about 35 nm, less than about 30 nm, less than about 25 nm, less than about 20 nm, less than about 15 nm, or less than 10 nm can be resolved using embodiments of the present invention.

One non-limiting example of such a method is stochastic optical reconstruction microscopy (STORM). See, e.g., U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, incorporated herein by reference, for examples of STORM techniques. In STORM, incident light is applied to emissive entities within a sample in a sample region to activate the entities, where the incident light has an intensity and/or frequency that is able to cause a statistical subset of the plurality of emissive entities to become activated from a state not capable of emitting light (e.g., at a specific wavelength) to a state capable of emitting light (e.g., at that wavelength). Once activated, the emissive entities may spontaneously emit light, and/or excitation light may be applied to the activated emissive entities to cause these entities to emit light. The excitation light may be of the same or different wavelength as the activation light. The emitted light can be collected or acquired, e.g., in one, two, or more objectives as previously discussed. In some cases, the excitation light is also able to subsequently deactivate the statistical subset of the

plurality of emissive entities, and/or the entities may be deactivated via other suitable techniques (e.g., by applying deactivation light, by applying heat, by waiting a suitable period of time, etc.). This process repeated as needed, each time with a statistically different subset of the plurality of emissive entities to emit light. In this way, a stochastic image of some or all of the emissive entities within a sample may be produced. In addition, as discussed herein, various image processing techniques, such as noise reduction and/or x, y and/or z position determination can be performed on the acquired images.

In some cases, incident light having a sufficiently weak intensity may be applied to a plurality of entities such that only a subset or fraction of the entities within the incident light are activated, e.g., on a stochastic or random basis. The amount of activation can be any suitable fraction, e.g., less than about 0.01%, less than about 0.03%, less than about 0.05%, less than about 0.1%, less than about 0.3%, less than about 0.5%, less than about 1%, less than about 3%, less than about 5%, less than about 10%, less than less than about 15%, less than about 20%, less than about 25%, less than about 30%, less than about 35%, less than about 40%, less than about 45%, less than about 50%, less than about 55%, less than about 60%, less than about 65%, less than about 70%, less than about 75%, less than about 80%, less than about 85%, less than about 90%, or less than about 95% of the entities may be activated, depending on the application. For example, by appropriately choosing the intensity of the incident light, a sparse subset of the entities may be activated such that at least some of them are optically resolvable from each other and their positions can be determined. In some embodiments, the activation of the subset of the entities can be synchronized by applying a short duration of incident light. Iterative activation cycles may allow the positions of all of the entities, or a substantial fraction of the entities, to be determined. In some cases, an image with sub-diffraction limit resolution can be constructed using this information.

Multiple locations on a sample can each be analyzed to determine the entities within those locations. For example, a sample may contain a plurality of various entities, some of which are at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light. Different locations within the sample may be determined (e.g., as different pixels within an image), and each of those locations independently analyzed to determine the entity or

entities present within those locations. In some cases, the entities within each location are determined to resolutions that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light, as previously discussed.

The emissive entities may be any entity able to emit light. For instance, the entity may be a single molecule. Non-limiting examples of emissive entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, fluorescent dyes such as cyanine dyes (e.g., Cy2, Cy3, Cy5, Cy5.5, Cy7, etc.), metal nanoparticles, semiconductor nanoparticles or "quantum dots," or fluorescent proteins such as GFP (Green Fluorescent Protein). Other light-emissive entities are known to those of ordinary skill in the art. As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 380 nm and about 750 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 700 micrometers and 1000 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one type of entity may be used, e.g., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities are chemically identical or at least substantially chemically identical.

In one set of embodiments, an emissive entity in a sample is an entity such as an activatable entity, a switchable entity, a photoactivatable entity, or a photoswitchable entity. Examples of such entities are discussed herein. In some cases, more than one type of emissive entity may be present in a sample. An entity is "activatable" if it can be activated from a state not capable of emitting light (e.g., at a specific wavelength) to a state capable of emitting light (e.g., at that wavelength). The entity may or may not be able to be deactivated, e.g., by using deactivation light or other techniques for deactivating light. An entity is "switchable" if it can be switched between two or more different states, one of which is capable of emitting light (e.g., at a specific wavelength). In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity can be "activated" to a first state able to produce light having a desired wavelength, and "deactivated" to a second state not able to produce light of the same wavelength.

If the entity is activatable using light, then the entity is a "photoactivatable" entity. Similarly, if the entity is switchable using light in combination or not in combination with other techniques, then the entity is a "photoswitchable" entity. For instance, a photoswitchable entity may be switched between different light-emitting or non-emitting states by incident light of different wavelengths. Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state, can no longer emit light (or emits light at a reduced intensity) or emits light at a different wavelength when exposed to the excitation wavelength. Examples of switchable entities are discussed below, and are also discussed in U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, incorporated herein by reference.

In one set of embodiments, a switchable entity may be used. Non-limiting examples of switchable entities (including photoswitchable entities) are discussed in U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, incorporated herein by reference. As a non-limiting example of a switchable entity, Cy5 can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, e.g., 633 nm, 647 nm or 657 nm red light can switch or deactivate Cy5 to a stable dark state, while 405 nm or 532 nm green light can switch or activate the Cy5 back to the fluorescent state. Other non-limiting examples of switchable entities include fluorescent proteins or inorganic particles, e.g., as discussed herein. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimulus (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimulus (e.g., a second wavelength of light or light with the first wavelength) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to be able to emit light, and the entity can

then emit light when excited by an excitation light. Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light.

In some cases, the activation light and deactivation light have the same wavelength. In some cases, the activation light and deactivation light have different wavelengths. In some cases, the activation light and excitation light have the same wavelength. In some cases, the activation light and excitation light have different wavelengths. In some cases, the excitation light and deactivation light have the same wavelength. In some cases, the excitation light and deactivation light have different wavelengths. In some cases, the activation light, excitation light and deactivation light all have the same wavelength.

The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric fields and/or magnetic fields. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

In one set of embodiments, the switchable entity can be immobilized, e.g., covalently, with respect to a binding partner, i.e., a molecule that can undergo binding with a particular analyte. Binding partners include specific, semi-specific, and non-specific binding partners as known to those of ordinary skill in the art. The term "specifically binds," when referring to a binding partner (e.g., protein, nucleic acid, antibody, etc.), refers to a reaction that is determinative of the presence and/or identity of one or other member of the binding pair in a mixture of heterogeneous molecules (e.g., proteins and other biologics). Thus, for example, in the case of a receptor/ligand binding pair, the ligand would specifically and/or preferentially select its receptor from a complex mixture of molecules, or vice versa. Other examples include, but are not limited to, an enzyme would specifically bind to its substrate, a nucleic acid would specifically bind to its complement, an antibody would specifically bind to its antigen. The binding may be by one or more of a variety of mechanisms including, but not

limited to ionic interactions, and/or covalent interactions, and/or hydrophobic interactions, and/or van der Waals interactions, etc. By immobilizing a switchable entity with respect to the binding partner of a target molecule or structure (e.g., DNA or a protein within a cell), the switchable entity can be used for various determination or imaging purposes. For example, a switchable entity having an amine-reactive group may be reacted with a binding partner comprising amines, for example, antibodies, proteins or enzymes.

In some embodiments, more than one switchable entity may be used, and the entities may be the same or different. In some cases, the light emitted by a first entity and the light emitted by a second entity have the same wavelength. The entities may be activated at different times and the light from each entity may be determined separately. This allows the location of the two entities to be determined separately and, in some cases, the two entities may be spatially resolved, even at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light (i.e., "sub-diffraction limit" resolutions). In certain instances, the light emitted by a first entity and the light emitted by a second entity have different wavelengths (for example, if the first entity and the second entity are chemically different, and/or are located in different environments). The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities or below the diffraction limit of the emitted light. In certain instances, the light emitted by a first entity and the light emitted by a second entity have substantially the same wavelengths, but the two entities may be activated by light of different wavelengths and the light from each entity may be determined separately. The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities, or below the diffraction limit of the emitted light.

In some cases, the entities may be independently switchable, i.e., the first entity may be activated to emit light without activating a second entity. For example, if the entities are different, the methods of activating each of the first and second entities may be different (e.g., the entities may each be activated using incident light of different wavelengths). As another non-limiting example, if the entities are substantially the same, a sufficiently weak intensity of light may be applied to the entities such that only a

subset or fraction of the entities within the incident light are activated, i.e., on a stochastic or random basis. Specific intensities for activation can be determined by those of ordinary skill in the art using no more than routine skill. By appropriately choosing the intensity of the incident light, the first entity may be activated without activating the second entity. The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities, or below the diffraction limit of the emitted light. As another non-limiting example, the sample to be imaged may comprise a plurality of entities, some of which are substantially identical and some of which are substantially different. In this case, one or more of the above methods may be applied to independently switch the entities. The entities may be spatially resolved even at distances of separation that are less than the wavelength of the light emitted by the entities, or below the diffraction limit of the emitted light.

In some embodiments, a microscope may be configured so to collect light emitted by the switchable entities while minimizing light from other sources of fluorescence (e.g., "background noise"). In certain cases, imaging geometry such as, but not limited to, a total-internal-reflection geometry, a spinning-disc confocal geometry, a scanning confocal geometry, an epi-fluorescence geometry, an epi-fluorescence geometry with an oblique incidence angle, etc., may be used for sample excitation. In some embodiments, as previously discussed, a thin layer or plane of the sample is exposed to excitation light, which may reduce excitation of fluorescence outside of the sample plane. A high numerical aperture lens may be used to gather the light emitted by the sample. The light may be processed, for example, using filters to remove excitation light, resulting in the collection of emission light from the sample. In some cases, the magnification factor at which the image is collected can be optimized, for example, when the edge length of each pixel of the image corresponds to the length of a standard deviation of a diffraction limited spot in the image.

In some embodiments of the invention, the switchable entities may also be resolved as a function of time. For example, two or more entities may be observed at various time points to determine a time-varying process, for example, a chemical reaction, cell behavior, binding of a protein or enzyme, etc. Thus, in one embodiment, the positions of two or more entities may be determined at a first point of time (e.g., as described herein), and at any number of subsequent points of time. As a specific

example, if two or more entities are immobilized relative to a common entity, the common entity may then be determined as a function of time, for example, time-varying processes such as movement of the common entity, structural and/or configurational changes of the common entity, reactions involving the common entity, or the like. The time-resolved imaging may be facilitated in some cases since a switchable entity can be switched for multiple cycles, with each cycle giving one data point of the position of the entity.

In some cases, one or more light sources may be time-modulated (e.g., by shutters, acoustic optical modulators, or the like). Thus, a light source may be one that is activatable and deactivatable in a programmed or a periodic fashion. In one embodiment, more than one light source may be used, e.g., which may be used to illuminate a sample with different wavelengths or colors. For instance, the light sources may emanate light at different frequencies, and/or color-filtering devices, such as optical filters or the like, may be used to modify light coming from the light sources such that different wavelengths or colors illuminate a sample.

Various image-processing techniques may also be used to facilitate determination of the entities. For example, drift correction or noise filters may be used. Generally, in drift correction, a fixed point is identified (for instance, as a fiduciary marker, e.g., a fluorescent particle may be immobilized to a substrate), and movements of the fixed point (i.e., due to mechanical drift) are used to correct the determined positions of the switchable entities. In another example method for drift correction, the correlation function between images acquired in different imaging frames or activation frames can be calculated and used for drift correction. In some embodiments, the drift may be less than about 1000 nm/min, less than about 500 nm/min, less than about 300 nm/min, less than about 100 nm/min, less than about 50 nm/min, less than about 30 nm/min, less than about 20 nm/min, less than about 10 nm/min, or less than 5 nm/min. Such drift may be achieved, for example, in a microscope having a translation stage mounted for *x*-*y* positioning of the sample slide with respect to the microscope objective. The slide may be immobilized with respect to the translation stage using a suitable restraining mechanism, for example, spring loaded clips. In addition, a buffer layer may be mounted between the stage and the microscope slide. The buffer layer may further restrain drift of the slide with respect to the translation stage, for example, by preventing

slippage of the slide in some fashion. The buffer layer, in one embodiment, is a rubber or polymeric film, for instance, a silicone rubber film.

Accordingly, one embodiment of the invention is directed to a device comprising a translation stage, a restraining mechanism (e.g., a spring loaded clip) attached to the translation stage able to immobilize a slide, and optionally, a buffer layer (e.g., a silicone rubber film) positioned such that a slide restrained by the restraining mechanism contacts the buffer layer. To stabilize the microscope focus during data acquisition, a "focus lock" device may be used in some cases. As a non-limiting example, to achieve focus lock, a laser beam may be reflected from the substrate holding the sample and the reflected light may be directed onto a position-sensitive detector, for example, a quadrant photodiode. In some cases, the position of the reflected laser, which may be sensitive to the distance between the substrate and the objective, may be fed back to a *z*-positioning stage, for example a piezoelectric stage, to correct for focus drift.

Another aspect of the invention is directed to a computer-implemented method. For instance, a computer and/or an automated system may be provided that is able to automatically and/or repetitively perform any of the methods described herein. As used herein, "automated" devices refer to devices that are able to operate without human direction, i.e., an automated device can perform a function during a period of time after a human has finished taking any action to promote the function, e.g., by entering instructions into a computer. Typically, automated equipment can perform repetitive functions after this point in time. The processing steps may also be recorded onto a machine-readable medium in some cases.

In some cases, a computer may be used to control excitation of the switchable entities and the acquisition of images of the switchable entities. In one set of embodiments, a sample may be excited using light having various wavelengths and/or intensities, and the sequence of the wavelengths of light used to excite the sample may be correlated, using a computer, to the images acquired of the sample containing the switchable entities. For instance, the computer may apply light having various wavelengths and/or intensities to a sample to yield different average numbers of activated switchable elements in each region of interest (e.g., one activated entity per location, two activated entities per location, etc.). In some cases, this information may

be used to construct an image of the switchable entities, in some cases at sub-diffraction limit resolutions, as noted above.

Still other embodiments of the invention are generally directed to a system able to perform one or more of the embodiments described herein. For example, the system may include a microscope, a device for activating and/or switching the entities to produce light having a desired wavelength (e.g., a laser or other light source), a device for determining the light emitted by the entities (e.g., a camera, which may include color-filtering devices, such as optical filters), and a computer for determining the spatial positions of the two or more entities.

In other aspects of the invention, the systems and methods described herein may also be combined with other imaging techniques known to those of ordinary skill in the art, such as high-resolution fluorescence *in situ* hybridization (FISH) or immunofluorescence imaging, live cell imaging, confocal imaging, epi-fluorescence imaging, total internal reflection fluorescence imaging, etc.

The following documents are incorporated herein by reference in their entireties: International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*, published as WO 2009/085218 on July 9, 2009; U.S. Patent No. 7,776,613, issued August 17, 2010, entitled "Sub-Diffraction Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; and U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.* Also incorporated herein by reference in its entirety is U.S. Provisional Patent Application Serial No. 61/576,089, filed December 15, 2011, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*

The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

EXAMPLE 1

Recent advances in super-resolution fluorescence microscopy have substantially improved the spatial resolution of optical imaging. The enhanced resolution has enabled the visualization of various cellular ultrastructures previously inaccessible to optical methods. Nonetheless, the current state-of-the-art resolution still cannot resolve many cellular structures, and further improvement of the resolution is desirable.

For example, among the three major types of cytoskeletal structures, individual microtubules and intermediate filaments can be routinely observed in cells with optical microscopy; by contrast, individual actin filaments have not been resolved in cells by optical techniques, including super-resolution methods, due to the small diameter and high packing density of actin filaments. Actin is of vital importance to many cellular processes. In particular, the assembly and disassembly of actin filaments in the thin sheet-like cell protrusions drive cell locomotion. The knowledge of how actin filaments are spatially organized is important for understanding these processes. While electron microscopy (EM) and cryo-electron tomography can resolve individual actin filaments in cells, three-dimensional (3D) reconstructions are still relatively challenging due in part to the structural perturbations induced by the dehydration and embedding treatment required for conventional EM and the difficulty in reconstructing large volumes by cryo-tomography. Therefore, except for the several hundred nanometers near the cell edge, a full 3D reconstruction of actin has not been achieved for the sheet-like cell protrusion, and how actin is vertically organized in this region is still unclear.

This example illustrates super-resolution fluorescence microscopy by combining 3D stochastic optical reconstruction microscopy (STORM) with a dual-objective detection scheme (Fig. 1B). In one embodiment of the 3D STORM technique, an optically resolvable subset of fluorescent probes are activated at any given instant, and their signals are detected using a single objective. Astigmatism is introduced in the detection path using a cylindrical lens such that the images obtained for individual molecules are elongated in x and y directions for molecules on the proximal and distal sides of the focal plane, respectively. The lateral and axial coordinates of the molecules are determined from the centroid positions and ellipticities of these single-molecule images, respectively. Iteration of the activation and imaging cycles allows the positions of numerous molecules to be determined and a super-resolution image to be reconstructed from these molecular coordinates. For additional details, see International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*, published as WO 2009/085218 on July 9, 2009, incorporated herein by reference in its entirety.

The 3D STORM method can be combined with a two-objective detection scheme to increase the image resolution of super-resolution fluorescence microscopy. In Fig.

1B, two microscope objectives are placed opposing each other and aligned so that they focus on the same spot of the sample. The sample, sandwiched between the two objectives, is illuminated with 647 nm and 405 nm lasers (using an optical fiber) through one of the objectives, and the fluorescence emission is collected by both objectives and projected onto two different areas of a single CCD camera. Astigmatism is introduced into the imaging path of both objectives using a cylindrical lens.

It may be reasoned that the total collected fluorescence signal would double by sandwiching the sample between two opposing objectives and detecting the signal from both objectives. Since the localization uncertainty of each molecule scales with the inverse square root of the number of photons detected, doubling the photon count would be expected to yield, at best, a 1.4-fold improvement in the image resolution. However, as discussed below, much higher resolutions were actually obtained.

To characterize the localization precision of the dual-objective STORM setup, individual Alexa 647 molecules scattered in fixed cells within ~150 nm of the focal plane were imaged. This range is generally comparable to the thickness of the sheet-like cell protrusions. While in these experiments, the imaging z-range was limited to be comparable to the thickness of cell protrusions, the imaging depth could be readily increased by stepping the sample in the z-direction. As each Alexa 647 molecule can be switched on and off multiple times, the standard deviation (SD) of repetitive localizations of the same molecule allowed an experimental determination of the localization precision.

Surprisingly, the measured localization precisions, ~4 nm and ~8 nm in x-y and z directions, respectively (Fig. 1C), represent a greater than two-fold improvement over previously reported values with the same fluorophore using one objective using 3D STORM techniques. This localization precision corresponds to an image resolution of ~9 nm in the lateral directions and ~19 nm in the axial direction, measured in full width at half maximum (FWHM).

Fig. 1C shows the localization precision of Alexa 647 molecules in fixed cells measured with this dual-objective system. Each molecule gives a cluster of localizations due to repetitive activation of the same molecule. Localizations from 108 clusters (each containing >10 localizations) are aligned by their center of mass to generate the 3D presentation of the localization distribution. Histograms of the distribution in x, y, and z

are fit to Gaussian functions, and the resultant standard deviations (σ_x, σ_y, and σ_z) are given in the plots.

The measured localization precision and image resolution are significantly higher than would have been predicted solely by doubling the photon collection efficiency with the use of two objectives, as discussed above. Indeed, at these length scales—less than 10 nm—any further significant improvements in resolution is both unexpected and important.

Further examination of the single-molecule images showed that beyond the expected doubling of photon-collection efficiency (Fig. 1D), the combination of astigmatism imaging and dual-objective detection provided a noise-cancelling mechanism that further improved image precision (Fig. 1E). Fig. 1D shows the distribution of the number of photons detected for individual Alexa 647 molecules through both objectives (Ave=10,600) in comparison to that detected from a single objective (Ave=5,200).

Fig. 1E shows images of activated Alexa 647 molecules obtained from the two objectives in a single frame, which demonstrates the anticorrelated appearance of entities in this particular embodiment of the invention. The scale bar is 2 micrometers. Molecules that were axially closer to one of the objectives were necessarily farther from the opposing one, resulting in anticorrelated changes in the ellipticity detected by the two objectives (Fig. 1E, left and center arrows). Thus, an entity that appears elongated in *x* through one objective should appear elongated in *y* through the opposing objective. In contrast, noise (such as those caused by sample drift) and abnormalities (such as two nearby molecules with overlapping images that are misidentified as a single molecule) led to correlated changes (Fig. 1E, right arrows). This effect allowed noises to be cancelled by averaging the *z* measurements from the two channels, and abnormalities to be identified and rejected by examining the difference in the *z*-positions obtained from the two objectives. In addition, the mechanical stability of the dual-objective setup (with the optical axis of the objectives parallel to the optical table and the sample vertical oriented) was found to be higher than that of the single-objective setup built on a standard inverted microscope with the optical axis of the objectives perpendicular to the optical table. These effects resulted in a substantial improvement of the spatial resolution.

EXAMPLE 2

This example illustrates imaging of an actin cytoskeleton using a configuration similar to the one discussed in Example 1. To benefit from high image resolution, the target structure was labeled using small organic molecules, by staining the actin filaments with Alexa 647 dye labeled phalloidin, which binds actin filaments with high specificity. Imaging using certain STORM techniques was performed through the direct activation of Alexa 647 using short-wavelength light. See, e.g., U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, or International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*, published as WO 2009/085218 on July 9, 2009, each incorporated herein by reference.

Fig. 2 compares the conventional and dual-objective STORM images of actin in a fibroblast (COS-7) cell. Fig. 2A shows a dual-objective image of actin (labeled with Alexa 647-phalloidin) in a COS-7 cell. The z-positions are shown using shading. The scale bar is 2 micrometers. Figs. 2B, 2C and 2D illustrate a close-up comparison between the dual-objective STORM image (Fig. 2B) with a single-objective STORM image (Fig. 2C) and a conventional fluorescence image (Fig. 2D) of the boxed region in Fig. 2A. The scale bar in these figures is 500 nm. Fig. 2E shows a cross-sectional profile of eight filaments overlaid by the center of each filament. The smooth line is a Gaussian fit with FWHM of 12 nm. Fig. 2F shows a cross-sectional profile for two nearby filaments obtained in the dual-objective image (identified in Figs. 2B and 2C by arrows) in comparison to the profile obtained in the single-objective image. The grey bars correspond to the dual-objective images in Fig. 2B and the line corresponds to the single objective image in Fig. 2C.

In contrast with the conventional fluorescence image, in which actin filaments were completely unresolvable (Fig. 2D), actin filaments were clearly resolved in dual-objective STORM images (Figs. 2B). The cross-sectional profile of individual filaments exhibited a 12 nm FWHM (Fig. 2E). After subtracting the effect of the 9-nm lateral image resolution, the width of the phalloidin-labeled actin filaments was found to be ~8 nm ($= (12^2 - 9^2)^{1/2}$), which agrees with the known diameter of actin (5 nm to 9 nm). In addition, nearby filaments separated by ~20 nm were well resolved from each other (Fig.

2F). In comparison, lower resolution was achieved if one only relies on the information collected by one of the two objectives (Figs. 2C and 2F)

The volumetric imaging capability with the 9-nm lateral and 19-nm axial resolutions further allowed a holistic, 3D view of the actin networks to be obtained. Referring now to Fig. 3, Fig. 3A shows a dual-objective image of actin in an epithelial (BSC-1) cell. The *z*-positions are shown as shading. Figs. 3B and 3C show vertical cross sections (each 500 nm wide in *x* or *y*) of the cell in Fig. 3A, along the dot and dash lines, respectively. Note when far from the cell edge, the *z*-position of the dorsal layer increases quickly and falls out of the imaging range used in this example. Figs. 3D and 3E show *z*-profiles for two points along the vertical section, corresponding to the left and right arrows in Fig. 3B, respectively. Each histogram is fit to two Gaussians (curves), yielding the apparent thickness of the ventral and dorsal layers and the peak separation between the two layers. Fig. 3F shows quantification of the apparent thickness averaged over the two layers and the dorsal-ventral separation obtained from the *x*-*z* cross-section profile in Fig. 3B. Figs. 3G and 3H show the ventral and dorsal actin layers of the cell in Fig. 3A. Figs. 3I and 3J show the ventral and dorsal actin layers of a COS-7 cell that was treated with blebbistatin. Figs. 3K and 3L show vertical cross sections (each 500 nm wide in *x* or *y*) of the blebbistatin-treated cell along the dot and dash lines, respectively. Fig. 3M shows the actin density of the ventral and dorsal layers along the horizontal boxes in Figs. 3I and 3J, measured by the localization density. The scale bars are 2 micrometers for Figs. 3A, 3G, 3H, 3I, and 3J, and 100 nm for *z* and 2 micrometers for *x* and *y* for Figs. 3B, 3C, 3K, and 3L.

In these figures, two vertically separated actin layers were observed in the sheet-like cell protrusion despite its small thickness (Figs. 3A-3C). Each of these layers was apparently about 30 to about 40 nm thick (Figs. 3D-3F). The separation between the two layers was generally ~100 nm, but could be as small as ~50 nm (Figs. 3D-3F). The separation increased to much larger than 200 nm in the interior region far from the cell edge, suggesting that the two layers evolve into the cortical actin layers in the cell body. The two layer organization was also validated in living cells.

Although filaments in the two layers formed well separated networks, thick filament bundles occasionally connected the two layers. Such bundles typically originated from adhesion plaques, ran through the ventral layer, and gradually rose

towards and ultimately reached the dorsal layer, as expected for the dorsal stress fibers. Thick bundles connecting adhesion plaques on the ventral surface, as expected for ventral stress fibers, were also observed. In addition to these mature focal adhesions at the end of the thick actin bundles, smaller and more isotropic adhesion complexes were also observed, likely representing nascent adhesions complexes. Actin filaments attached to these structures diverged in different directions and often connected nearby adhesion plaques.

Remarkably, the two layers of actin networks exhibited highly distinct spatial organizations of actin filaments (Figs. 3G and 3H). While the dorsal layer typically appeared as a consistently dense and homogeneous meshwork, the ventral layer formed a web-like structure with a lower filament density and highly variable organization. The two-layer arrangement was consistently observed in all BSC-1 epithelial cells that were imaged as well as in COS-7 fibroblast cells. The actin density in the dorsal layer could be several times as high as that in the ventral layer. Additional analysis suggests that the two layer arrangement spans the lamellum and possibly extends into the lamellipodium.

To explore the molecular mechanisms underlying the structural differences observed for the two actin networks, how these networks responded to different actomyosin-perturbing drugs was investigated. Cytochalasin D, a drug that inhibits actin polymerization, reduced the filament density as expected. Latrunculin A, a drug that sequesters monomeric actin, also reduced the filament density as expected, but with the dorsal network substantially more disrupted than the ventral layer, suggesting that the dorsal layer is potentially more dynamic and thus more readily disrupted by actin monomer depletion. Interestingly, blebbistatin, an inhibitor for myosin II, completely removed the structural differences between the ventral and dorsal networks: both networks became uniform actin meshworks of similar density, reminiscent of the dense, uniform dorsal network observed in untreated cells (Figs. 3I-M). These results suggest that myosin II plays a key role in the structural reorganization of the ventral actin layer and in maintaining the structural differences between the dorsal and ventral actin networks. This function of myosin II is potentially related to its activities previously found in regulating actin disassembly, actin bundle formation, and focal adhesion maturation.

In summary, by combining astigmatism imaging, dual-objective detection and small-molecule labeling in STORM, image resolutions of <10 nm in the lateral directions and <20 nm in the axial direction may be obtained. This lateral resolution is 2-fold or more higher than previously achieved for biological samples using other super-resolution methods. Although higher axial resolutions have been reported using interferometry approaches, the astigmatism-based system reported here is substantially simpler to implement.

With the improved resolution, these examples illustrate the resolution of individual actin filaments in cells for the first time using fluorescence microscopy, which opens a new window for studying numerous actin-related processes in cells. These examples also illustrate the 3D ultrastructure of the actin cytoskeleton in sheet-like cell protrusions and revealed two layers of continuous actin networks with distinct structures, which both supports and extends previous understandings. In addition to these examples, the high image resolution obtained with dual-objective STORM should also find use in many other systems.

EXAMPLE 3

This example describes the optical setup used in Examples 1 and 2. A schematic of the dual-objective setup is shown in Fig. 1B. Two infinity-corrected microscope objectives (Olympus Super Apochromat UPLSAPO 100x, oil immersion, NA 1.40) were placed opposing each other and aligned so they focus on the same spot of the sample. A piezoelectric actuator (Thorlabs DRV120) was used to control the axial position of the sample with nanometer precision. The 647 nm line from a Kr/Ar mixed gas laser (Innova 70C Spectrum, Coherent) and the 405 nm beam from a solid state laser (CUBE 405-50C, Coherent) were introduced into the sample through the back focal plane of the first objective using a customized dichroic mirror that worked at an incident angle of 22.5^{o} (Chroma). A translation stage allowed the laser beams to be shifted towards the edge of the objective so that the emerging light reached the sample at incidence angles slightly smaller than the critical angle of the glass-water interface, thus illuminating only the fluorophores within a few micrometers of the coverslip surface. The fluorescence emission was collected by both objectives. After passing through long-pass filters (HQ665LP, Chroma), the two parallel light rays from the two objectives were each focused by a 20 cm achromatic lens, cropped by a slit at the focal plane, and then

separately projected onto two different areas of the same EMCCD camera (Andor iXon DU-897) using two pairs of relay lenses. Astigmatism was introduced into the imaging paths of both objectives using a cylindrical lens so that the images obtained by each objective were elongated in *x* and *y* for molecules on the proximal and distal sides of the focal plane (relative to the objective), respectively. A band-pass filter (ET700/75m, Chroma) was installed on the camera.

Following is one example of a procedure for setting up the optical setup discussed above.

Combine different laser lines using dichroic mirrors and/or prisms. For studies using Alexa 647 dye labels, a red laser (e.g., the 647 nm line from a Kr/Ar mixed gas laser or a 656 nm solid state laser) and a violet or UV laser (e.g., a 405 nm solid state laser) can be used. An AOTF (acousto-optical tunable filter) may be used to control both the shuttering and the light intensity for different laser lines. Couple the combined laser lights into the optical fiber.

Combine the 2D translation stage with a 1D translation stage to obtain 3D control of the sample position. For actuation of the 1D translation stage (e.g., for *z* positioning of the sample), combine a DRV120 piezoelectric actuator with a DRV3 manual actuator to obtain both a large (8 mm) working distance for coarse alignment (DRV3) and nanometer precision (e.g., within a range of ~20 micrometers) for fine adjustments (DRV120). Center the sample stage for initial alignment.

Mount Objective 1 ("Obj. 1") using a *z*-axis translation mount. Use a calibration slide as the sample and illuminate from the opposite side with white light (this can be easily done as Objective 2 ("Obj. 2") has not been installed yet at this step). Add in the 22.5° dichroic mirror and the tube lens (L3) for Objective 1, and place Slit 1 at the intermediate image formed after L3 and M3. Open the slit. Project the image onto the EMCCD camera using a pair of relay lenses (L4 and L5). Align the camera so that the center of the image is projected onto the center of the right half of the CCD. Use Slit 1 to crop the image so that when looking at the acquired camera signal on the computer screen, the image is restrained to only one half of the camera.

Collimate the laser light coming out of the end of the optical fiber and introduce it into the sample through the back focal plane of Objective 1. Place M1, M2, and L2 on a 1D translation stage so the laser beam can be shifted to the edge of the objective during

imaging. During initial alignment, however, pass the laser beam through the center of Objective 1 and make sure that the laser is parallel to the axis of the objective by adjusting M1 and M2. Multiple irises can be placed along the light path to check the alignment.

Mount Objective 2 with an *x-y* translator. Illuminate the calibration slide through Objective 1 using a weak laser and add in M5, L6, Slit 2, L7, M6, and L8 sequentially, similar to that was done for Objective 1, but project the final image onto the left half of the CCD instead.

Insert the two long-pass filters (LP1 and LP2) into the two optical paths. Add in the cylindrical lens after the relay lenses (L5 and L8) of the optical paths for both Objective 1 and Objective 2. Install the band-pass filter on the camera.

Assemble a fluorescent bead sample with coverslips on both sides. Replace the calibration slide with the bead sample. Use the 647 nm laser to illuminate the sample. Use the DRV120-DRV3 actuator system (Step 2) to adjust the *z*-position of the sample so that the beads are in focus for Objective 1. Align Objective 2 by adjusting the *x-y* translator and the *z*-axis translation mount so that Objective 1 and Objective 2 focus on the same spot of the bead sample. Make fine adjustments to Slit 1 and Slit 2 while observing the acquired camera signal on the computer screen, to ensure that the two images obtained via Objective 1 and Objective 2 each occupies one-half of the CCD without overlapping.

For single molecule imaging, background light should be reduced or eliminated. For example, a box can be built around the camera and the relay lenses, as illustrated in the schematic diagram in Fig. 1B.

To enhance image contrast, adjust the translation stage for the incoming laser beam to shift it to the edge of Objective 1, so that the emerging light reaches the sample at incidence angles slightly smaller than the critical angle of the glass-water interface, thus illuminating only the fluorophores within a few micrometers of the coverslip surface.

For imaging of the Alexa 647 labeled actin, use 647 nm laser (~2 kW/cm^2) to excite fluorescence from Alexa 647 molecules and switch them into the dark state. Use the 405 nm laser to reactivate the fluorophores from the dark state back to the emitting state. Adjust the power of the 405 nm laser (0 W/cm^2 to 1 W/cm^2) during image

acquisition so that at any given instant, only a small, optically resolvable fraction of the fluorophores in the sample were in the emitting state. See, e.g., U.S. Patent No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*, or International Patent Application No. PCT/US2008/013915, filed December 19, 2008, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*, published as WO 2009/085218 on July 9, 2009, each incorporated herein by reference, for more details of STORM imaging.

It should be noted that the above discussion is only one example of a procedure for setting up the optical setup shown in Fig. 1B, and other techniques may also be used. In addition, other optical setups may also be used in other embodiments of the invention besides the one illustrated in Fig. 1B.

EXAMPLE 4

Following are various protocols useful in the examples discussed above.

Sample preparation. BSC-1 and COS-7 cells were plated on 18-mm diameter, #1.5 uncoated glass coverslips at a confluency of ~20%. After 16-24 hours, the cells were fixed and labeled following previously developed protocols for ultrastructural studies of actin cytoskeleton. Briefly, the cells were initially fixed and extracted for 1-2 min using a solution of 0.3% glutaraldehyde and 0.25% Triton X-100 in cytoskeleton buffer (CB: 10 mM MES pH 6.1, 150 mM NaCl, 5 mM EGTA, 5 mM glucose, and 5 mM $MgCl_2$), and then post-fixed for 10 min in 2% glutaraldehyde in CB. The sample was treated with freshly-prepared 0.1% sodium borohydride for 7 min to reduce background fluorescence. For vinculin staining (when needed), the sample was first blocked with 3% BSA and 0.5% Triton X-100, and then stained with rabbit monoclonal vinculin antibodies (Invitrogen 700062) followed by Cy3-labled goat anti-rabbit secondary antibodies (Invitrogen A10520). Actin filaments were labeled with Alexa Fluor 647-phalloidin (Invitrogen A22287) overnight at 4 ^{o}C. A concentration of ~0.5 micromolar phalloidin in phosphate buffered saline (PBS) was used. To minimize the dissociation of phalloidin from actin, the sample was briefly washed once with PBS and then immediately mounted for STORM imaging.

For drug-effect studies, cells were incubated with culture media containing either 0.5 micromolar cytochalasin D (Sigma-Aldrich), 0.25 micromolar latrunculin A

(Invitrogen), or 50 micromolar (–)-blebbistatin (the active enantiomer; Sigma-Aldrich) at 37 °C for 1 hour, and then fixed and labeled as described above.

The imaging buffer for fixed cells was PBS with the addition of 100 mM cysteamine, 5% glucose, 0.8 mg/mL glucose oxidase (Sigma-Aldrich), and 40 micrograms/mL catalase (Roche Applied Science). ~4 microliters of imaging buffer was dropped at the center of a freshly-cleaned, #1.5 rectangular coverslip (22 mm by 60 mm), and the sample on the 18-mm diameter coverslip was mounted on the rectangular coverslip and sealed with nail polish.

Image data acquisition. The sealed sample was mounted between the two opposing objectives. The 647 nm laser was used to excite fluorescence from Alexa Fluor 647 molecules. Prior to acquiring images, a relatively weak 647 nm light (~0.05 W/cm^2) was used to illuminate the sample and recorded the conventional fluorescence image before any substantial fraction of the dye molecules were switched off. The 647 nm light intensity was then increased (to ~2 kW/cm^2) to rapidly switch the dyes off for STORM imaging. The 405 nm laser was used to reactivate the fluorophores from the dark state back to the emitting state. The power of the 405 nm laser (0-1 W/cm^2) was adjusted during image acquisition so that at any given instant, only a small, optically resolvable fraction of the fluorophores in the sample were in the emitting state. The EMCCD camera acquired images from both objectives simultaneously at a frame rate of 60 Hz. Typically, ~90,000 frames were recorded to generate the final super-resolution images. Recording of more frames (e.g., 230,000 frames for Fig. 2) further improved the image quality at the expense of longer imaging time.

Image data analysis. The recorded data were first split into two movies, each of which comprised a sequence of images obtained by one of the two objectives. Each movie was first analyzed separately according to previously described methods (see, e.g., U.S. Patent No. 7,838,302, incorporated herein by reference). The centroid positions and ellipticities of the single-molecule images provided lateral and axial positions of each activated fluorescent molecule, respectively. The molecular positions obtained by the second objective were mapped to the coordinates of the first objective through a transformation based on corresponding features (control points) in both images. The mapped data from the two objectives were then compared frame-by-frame: molecules that were switched on within one-frame of time and that were within ~50 nm to each

other in the mapped *x-y* plane were identified as the same emitting molecule detected by both objectives. Non-matching molecules were discarded. For each pair of matched molecules observed by the two objectives, the availability of two *z*-positions obtained through the two objectives provides a technique to identify abnormalities and cancel noise. Since the focal planes of the two opposing objectives coincided, a molecule on the side of the focal plane proximal to one objective would be on the distal side for the other objective. Therefore, its image would appear elongated in *x* through one objective but elongated in *y* through the other objective. A real change in the *z* position would cause anticorrelated changes in ellipticity measured through the two objectives (Fig. 1E, left and center arrows). On the other hand, abnormalities and noise would tend to cause correlated changes in ellipticity. For example, when two close-by molecules with overlapping images were misidentified as a single molecule, the resultant images through both objectives would appear elongated in the same direction along the line connecting the two molecules (Fig. 1E, right arrows). Likewise, any *x-y* drift of the stage or the camera would also cause elongation in the same direction. These correlated changes in ellipticity resulted in apparently different *z* positions obtained through the two objectives (delta-*z*). The value of delta-*z* (Δz) could thus be used to identify and reject abnormalities. These abnormalities (identified by delta-$z > 100$ nm, which is substantially larger than the axial resolution of a single objective) amounted to ~10% of all identified entities. For molecule pairs that matched well with each other in all four (spatial and temporal) coordinates, the final coordinates were determined as the average of the mapped coordinates from the two objectives, weighted by the width of the image and number of photons obtained by each objective. This averaging procedure further reduced noise caused by errors, such as the correlated changes in ellipticity described above. The final super-resolution images were reconstructed from these molecular coordinates by depicting each location as a 2D Gaussian peak.

To characterize the localization precision, fixed cell samples sparsely labeled with Alexa Fluor 647 were used. Relatively strong activation conditions were used such that each Alexa Fluor 647 molecule was activated multiple times during the image acquisition time and gave a cluster of localizations due to the repetitive activation. The sparse labeling condition allowed the identification of these clusters of localizations. Localizations from many such clusters within ~150 nm of the focal plane were aligned

by their center of mass to generate the localization distribution reported in Fig. 1C. The localization precision determined from these distributions was ~4 nm (SD) in the *x-y* directions and ~8 nm (SD) in the *z* direction, respectively. The variation in localization precision across this region was small (within 15% of the average value). The localization precision determined from the sparsely labeled sample is a good representation for the densely labeled actin samples as the parameters relevant for the localization precision, such as the number of photons detected from individual molecules and the background fluorescence signal, were measured to be the same for both sparsely and densely labeled samples.

For STORM imaging of the densely labeled actin samples, relatively weak 405 nm activation intensities were used during the image acquisition time of ~90,000 frames. This led to a typical linear localization density of 1 localization per 4 nm along individual actin filaments, which corresponds to a Nyquist Criterion based resolution of 8 nm, smaller than the 9 nm lateral and 19 nm axial resolutions determined above. As not all of the labeled molecules had been exhausted by the end of the 90,000 frames, the number of localizations could still be further increased by increasing the imaging time.

For the characterization of the widths of individual actin filaments, short (~200 nm), straight segments of filaments in cells were chosen, along which no crossing or branching of filaments were observed. Analysis of 10 such segments yielded FWHM widths of 11 +/- 2 nm.

While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention

described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will

refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

CLAIMS

1. A microscopy system, comprising:

a sample region;

a first objective on a first side of the sample region;

a second objective on a second side of the sample region; and

a non-circularly-symmetric lens positioned in a first imaging path in optical communication with the first objective.

2. The microscopy system of claim 1, wherein the first objective and the second objective are collinearly positioned relative to each other.

3. The microscopy system of any one of claims 1 or 2, wherein the first objective and the second objective are positioned at about 180° relative to each other.

4. The microscopy system of any one of claims 1-3, wherein the non-circularly-symmetric lens is cylindrical lens.

5. The microscopy system of any one of claims 1-4, wherein the non-circularly-symmetric lens is positioned in a second imaging path in optical communication with the second objective.

6. The microscopy system of any one of claims 1-5, further comprising a second non-circularly-symmetric lens positioned in a second imaging path in optical communication with the second objective.

7. The microscopy system of any one of claims 1-6, wherein the non-circularly-symmetric lens produces elliptical images of entities in the sample region.

8. The microscopy system of any one of claims 1-7, wherein the non-circularly-symmetric lens defines a focal region, wherein at least a portion of the sample does not overlap with the focal region.

9. The microscopy system of any one of claims 1-8, wherein the sample region is substantially vertically positioned.

10. The microscopy system of any one of claims 1-9, wherein the sample region is substantially planar.

11. The microscopy system of any one of claims 1-10, further comprising a detector in optical communication with the first objective via the first imaging path.

12. The microscopy system of claim 11, wherein the detector is further in optical communication with the second objective via a second imaging path.

13. The microscopy system of any one of claims 11 or 12, wherein the detector is a CCD camera.

14. The microscopy system of any one of claims 11-13, wherein at least some of the light from the sample region is not focused on the detector.

15. The microscopy system of any one of claims 11-14, further comprising a second detector in optical communication with the second objective via a second imaging path.

16. The microscopy system of any one of claims 1-15, further comprising an illumination path that intersects the sample region.

17. The microscopy system of claim 16, wherein the illumination path intersects the sample region with an incidence angle larger than about 55° relative to optical axis of the imaging system.

18. The microscopy system of any one of claims 16 or 17, wherein the illumination path intersects the sample region with an incidence angle that is smaller than the critical angle of a glass-water interface.

19. The microscopy system of claim 18, wherein the illumination path is directed through a dichroic mirror.

20. The microscopy system of any one of claims 1-19, further comprising a translation stage able to manipulate at least a portion of the sample region.

21. The microscopy system of claim 20, wherein the translation stage is piezoelectrically controlled.

22. The microscopy system of any one of claims 1-21, wherein the first imaging path is deflected via at least one mirror.

23. The microscopy system of any one of claims 1-22, wherein the first imaging path passes through at least one slit.

24. The microscopy system of any one of claims 1-23, further comprising a first illumination source in optical communication with at least a portion of the sample region.

25. The microscopy system of claim 24, wherein the first illumination source is substantially monochromatic.

26. The microscopy system of any one of claims 24 or 25, further comprising second illumination source in optical communication with at least a portion of the sample region.

27. The microscopy system of claim 26, wherein the second illumination source is substantially monochromatic.

28. The microscopy system of any one of claims 1-27, wherein the first objective has a magnification of at least 80x.

29. The microscopy system of any one of claims 1-28, wherein the first objective has a magnification of at least 100x.

30. A method, comprising:

acquiring a first plurality of images from a first side of a sample;

acquiring a second plurality of images from a second side of the sample; and

comparing the first and second plurality of images to determine positions of one or more entities in the sample by determining the shapes and/or intensities of the appearance of the entities present in the first and second plurality of images.

31. The method of claim 30, comprising acquiring the first plurality of images using a stochastic imaging technique.

32. The method of claim 31, comprising acquiring the first plurality of images using Stochastic Optical Reconstruction Microscopy (STORM).

33. The method of any one of claims 31 or 32, wherein at least some of the entities are emissive entities, and wherein the stochastic imaging technique used to acquire the first plurality of images comprises:

applying incident light to the sample, wherein the incident light is able to cause a statistical subset of the plurality of emissive entities to emit light, and to subsequently deactivate the statistical subset of the plurality of emissive entities;

acquiring the light emitted by the statistical subset of the plurality of emissive entities to produce an image; and

repeating the above two acts one or more times, each time causing a

statistically different subset of the emissive entities to emit light, thereby producing the first plurality of images.

34. The method of claim 33, wherein at least some of the plurality of emissive entities are fluorescent.

35. The method of any one of claims 33 or 34, wherein at least some of the plurality of emissive entities are photoactivatable.

36. The method of any one of claims 33-35, wherein at least some of the plurality of emissive entities are photoswitchable.

37. The method of any one of claims 30-36, wherein at least some of the entities are photoswitchable fluorescent entities.

38. The method of any one of claims 30-37, comprising determining *x*, *y*, and *z* positions of at least some of the entities.

39. The method of any one of claims 30-38, comprising determining the *x*-*y* positions of at least some of the entities to a resolution of smaller than about 500 nm.

40. The method of any one of claims 30-39, comprising determining the *x*-*y* positions of at least some of the entities to a resolution of smaller than about 100 nm.

41. The method of any one of claims 30-40, comprising determining the *x*-*y* positions of at least some of the entities to a resolution of smaller than about 50 nm.

42. The method of any one of claims 30-41, comprising determining the *x*-*y* positions of at least some of the entities to a resolution of smaller than about 20 nm.

43. The method of any one of claims 30-42, comprising determining the *x*-*y* positions of at least some of the entities to a resolution of smaller than about 10 nm.

44. The method of any one of claims 30-43, comprising determining the z position of at least some of the entities to a resolution of smaller less than about 1000 nm.

45. The method of any one of claims 30-44, comprising determining the z position of at least some of the entities to a resolution of smaller than about 500 nm.

46. The method of any one of claims 30-45, comprising determining the z position of at least some of the entities to a resolution of smaller than about 100 nm.

47. The method of any one of claims 30-46, comprising determining the z position of at least some of the entities to a resolution of smaller than about 50 nm.

48. The method of any one of claims 30-47, comprising determining the z position of at least some of the entities to a resolution of smaller less than about 10 20 nm.

49. The method of any one of claims 30-48, wherein at least some of the entities are emissive entities, the method comprising determining x, y, and z positions of the emissive entities to a resolution of less than the wavelengths of light emitted by the emissive entities.

50. The method of any one of claims 30-49, comprising determining the ellipticity of the appearance of at least some of the entities in the first and second plurality of images.

51. The method of any one of claims 30-50, comprising acquiring the first plurality of images by imaging the sample through a non-circularly-symmetric lens.

52. The method of claim 51, wherein the non-circularly-symmetric lens is a cylindrical lens.

53. The method of any one of claims 51 or 52, comprising acquiring the second plurality of images by imaging the sample through the non-circularly-symmetric lens.

54. The method of any one of claims 30-53, comprising rejecting an entity due to non-anticorrelated changes between the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images.

55. The method of any one of claims 30-54, comprising accepting an entity due to anticorrelated changes between the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images.

56. The method of any one of claims 30-55, further comprising assembling the first and second plurality of images into a final data set.

57. The method of any one of claims 30-56, further comprising assembling the first and second plurality of images into a final image.

58. The method of any one of claims 30-57, comprising constructing an image from the positions of the entities.

59. The method of claim 58, where in the image has sub-diffraction-limit resolution

60. The method of any one of claims 30-59, wherein the first plurality of images and the second plurality of images are images of the same region of the sample.

61. The method of any one of claims 30-60, wherein at least some images of the first plurality of images and at least some images of the second plurality of images are obtained substantially simultaneously.

62. The method of any one of claims 30-61, wherein the first plurality of images and the second plurality of images are obtained using a common detector.

63. The method of any one of claims 30-62, wherein the sample comprises a cell.

64. The method of any one of claims 30-63, wherein the sample is substantially vertically positioned.

65. A method, comprising:

acquiring a first plurality of images from a first side of a sample;

acquiring a second plurality of images from a second side of the sample; and

comparing the first and second plurality of images to determine positions of one or more entities in the sample to a resolution of less than about 1000 nm without using interference between light that forms the first plurality of images and light that forms the second plurality of images.

66. The method of claim 65, comprising determining the shapes and/or intensities of the appearance of the entities in the first and second plurality of images.

67. The method of claims 65 or 66, comprising determining the positions in a direction substantially normal to the first side of the sample.

68. The method of any one of claims 65-67, comprising determining the ellipticity of the appearance of at least some of the entities in the first and second plurality of images.

69. The method of any one of claims 65-68, comprising acquiring the first plurality of images using a stochastic imaging technique.

70. The method of claim 69, comprising acquiring the first plurality of images using Stochastic Optical Reconstruction Microscopy (STORM).

71. The method of any one of claims 69 or 70, wherein at least some of the entities are emissive entities, and wherein the stochastic imaging technique used to acquire the first plurality of images comprises:

applying incident light to the sample, wherein the incident light is able to cause a statistical subset of the plurality of emissive entities to emit light, and to subsequently deactivate the statistical subset of the plurality of emissive entities;

acquiring the light emitted by the statistical subset of the plurality of emissive entities to produce an image; and

repeating the above two acts one or more times, each time causing a statistically different subset of the emissive entities to emit light, thereby producing the first plurality of images.

72. The method of claim 71, wherein at least some of the plurality of emissive entities are fluorescent.

73. The method of any one of claims 71 or 72, wherein at least some of the plurality of emissive entities are photoactivatable.

74. The method of any one of claims 71-73, wherein at least some of the plurality of emissive entities are photoswitchable.

75. The method of any one of claims 65-74, comprising acquiring the first plurality of images by imaging the sample through a non-circularly-symmetric lens.

76. The method of any one of claims 65-75, comprising rejecting an entity due to non-anticorrelated changes between the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images.

77. The method of any one of claims 65-76, comprising accepting an entity due to anticorrelated changes between the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images.

78. The method of any one of claims 65-77, wherein at least some images of the first plurality of images and at least some images of the second plurality of images are obtained substantially simultaneously.

79. A method, comprising:

acquiring a first plurality of images from a first side of a sample;

acquiring a second plurality of images from a second side of the sample; and

comparing the first and second plurality of images to determine positions of one or more emissive entities in the sample, to a resolution of less than the wavelengths of the light emitted by the emissive entities, without using interference between light that forms the first plurality of images and light that forms the second plurality of images.

80. A method, comprising:

acquiring a first plurality of images from a first side of a sample by imaging the sample through a non-circularly-symmetric lens; and

acquiring a second plurality of images from a second side of the sample.

81. The method of claim 80, comprising acquiring the first plurality of images using a stochastic imaging technique.

82. The method of any one of claims 80 or 81, comprising acquiring the first plurality of images using Stochastic Optical Reconstruction Microscopy (STORM).

83. The method of any one of claims 80-82, comprising acquiring the second plurality of images by imaging the sample through the non-circularly-symmetric lens.

84. A method, comprising:

providing a sample comprising photoswitchable fluorescent entities;

activating a subset of the photoswitchable fluorescent entities;

acquiring a first plurality of images from a first side of the sample; and

acquiring a second plurality of images from a second side of the sample.

85. The method of claim 84, comprising acquiring the first plurality of images by imaging the sample through a non-circularly-symmetric lens.

86. A method, comprising:

providing a sample comprising photoswitchable fluorescent entities;

acquiring a first plurality of images from a first side of the sample using a stochastic imaging technique; and

acquiring a second plurality of images from a second side of the sample using the stochastic imaging technique.

87. The method of claim 86, comprising acquiring the first plurality of images by imaging the sample through a non-circularly-symmetric lens.

88. A method, comprising:

providing a sample comprising photoswitchable fluorescent entities;

acquiring a first plurality of images from a first side of the sample;

acquiring a second plurality of images from a second side of the sample; and

determining x, y, and z positions of at least one of the photoswitchable fluorescent entities in the sample, using the first and second plurality of images, to a resolution of less than about 1000 nm.

89. The method of claim 88, comprising acquiring the first plurality of images using a stochastic imaging technique.

90. The method of any one of claims 88 or 89, comprising acquiring the first plurality of images using Stochastic Optical Reconstruction Microscopy (STORM).

91. The method of any one of claims 88-90, comprising acquiring the second plurality of images by imaging the sample through the non-circularly-symmetric lens.

92. A method, comprising:

providing a sample comprising photoswitchable fluorescent entities;

acquiring a first plurality of images from a first side of the sample;

acquiring a second plurality of images from a second side of the sample; and

determining x, y, and z positions of at least one of the photoswitchable fluorescent entities in the sample, using the first and second plurality of images, to a resolution of less than a wavelength of light emitted by the photoswitchable fluorescent entity.

93. A method, comprising:

providing a sample comprising one or more entities;

acquiring a first plurality of images from a first side of the sample;

acquiring a second plurality of images from a second side of the sample;

accepting entities due to anticorrelated changes between the appearance of the entities in the first plurality of images and the appearance of the entities in the second plurality of images; and

assembling the accepted entities into a final data set or image.

94. The method of claim 93, further comprising rejecting entities due to non-anticorrelated changes between the entity in the first plurality of images and the entity in the second plurality of images.

95. The method of any one of claims 93 or 94, comprising acquiring the first plurality of images using a stochastic imaging technique.

96. The method of claim 95, comprising acquiring the first plurality of images using Stochastic Optical Reconstruction Microscopy (STORM).

97. The method of any one of claims 93-96, comprising acquiring the first plurality of images by imaging the sample through the non-circularly-symmetric lens.

98. The method of any one of claims 93-97, wherein the final data set comprises positions of the accepted entities.

99. A method, comprising:

providing a sample comprising one or more entities;

acquiring a first plurality of images from a first side of the sample;

acquiring a second plurality of images from a second side of the sample;

rejecting entities due to correlated changes between the appearance of the entity in the first plurality of images and the appearance of the entity in the second plurality of images; and

assembling the first and second plurality of images into a final data set or image while suppressing the rejected entities.

100. A microscopy system, comprising:

a substantially vertically-positioned sample region;

a first objective on a first side of the sample region; and

a second objective on a second side of the sample region.

101. The microscopy system of claim 100, further comprising a non-circularly-symmetric lens positioned in a first imaging path in optical communication with the first objective.

102. The microscopy system of any one of claims 100 or 101, wherein the first objective and the second objective are collinearly positioned relative to each other.

103. A microscopy system, comprising:

a sample region;

a first objective on a first side of the sample region;

a second objective on a second side of the sample region; and

means for acquiring a super-resolution image of a sample in the sample region.

104. The microscopy system of claim 103, wherein the first objective and the second objective are collinearly positioned relative to each other.

105. The microscopy system of any one of claims 103 or 104, further comprising a non-circularly-symmetric lens positioned in a first imaging path in optical communication with the first objective.

106. The microscopy system of any one of claims 103-105, wherein the sample region is substantially vertically positioned.

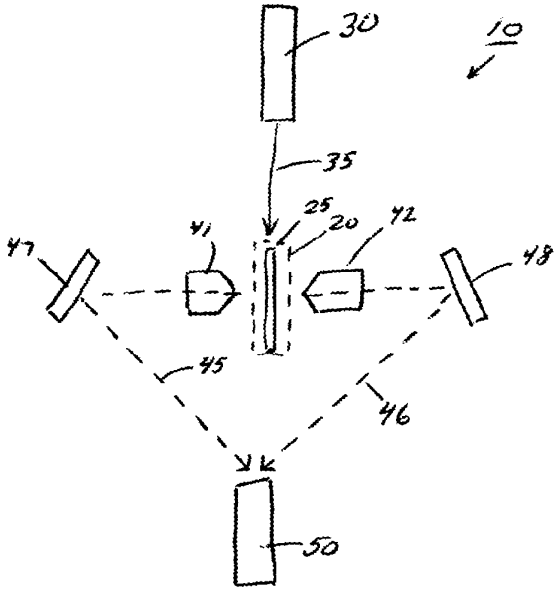

30
10
35
25
20
41
42
47
48
45
46
50

Fig. 1A

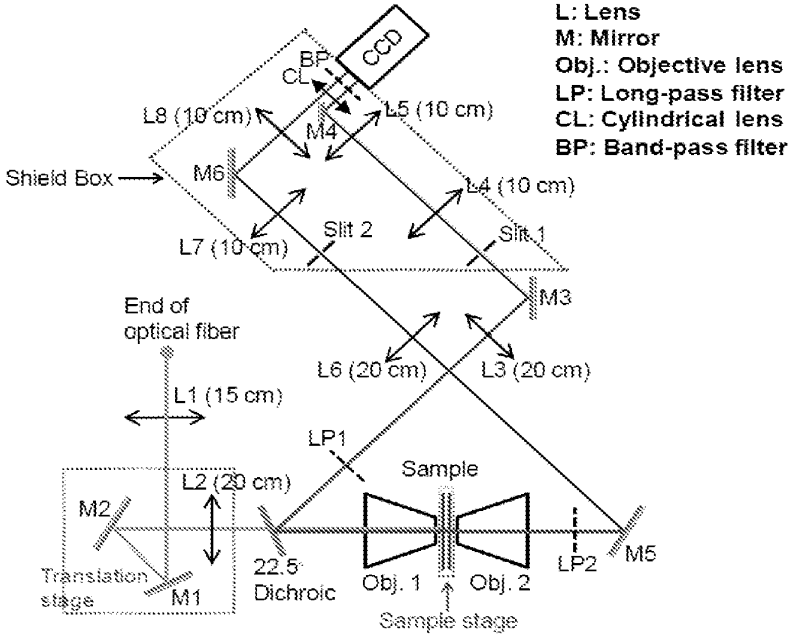

L: Lens
M: Mirror
Obj.: Objective lens
LP: Long-pass filter
CL: Cylindrical lens
BP: Band-pass filter
CCD
BP
CL
L8 (10 cm)
L5 (10 cm)
M4
Shield Box
M6
L4 (10 cm)
L7 (10 cm)
Slit 2
Slit 1
M3
End of optical fiber
L6 (20 cm)
L3 (20 cm)
L1 (15 cm)
LP1
Sample
L2 (20 cm)
M2
Translation stage
M1
22.5° Dichroic
Obj. 1
Obj. 2
LP2
M5
Sample stage

Fig. 1B

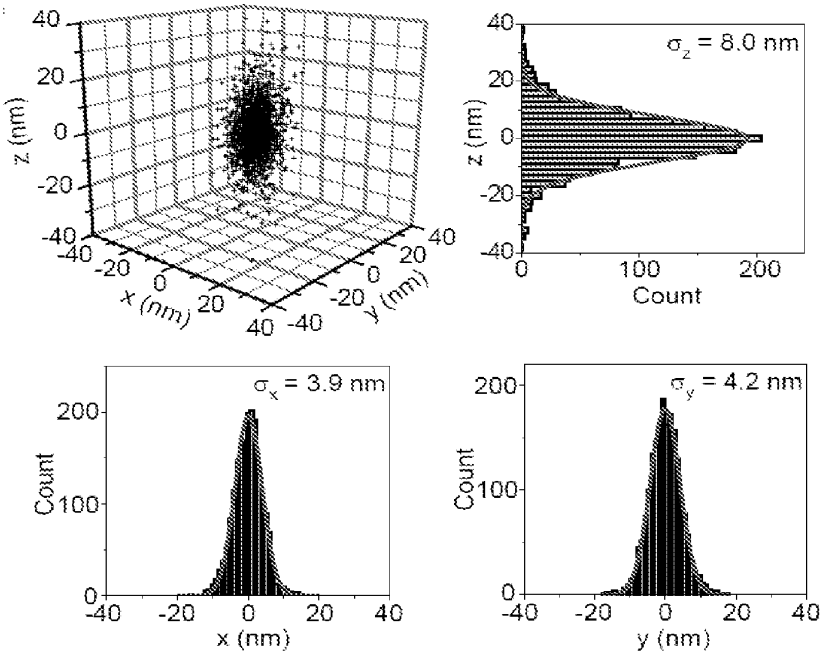

z (nm)
x (nm)
y (nm)
σz = 8.0 nm
Count
σx = 3.9 nm
σy = 4.2 nm
x (nm)
y (nm)

Fig. 1C

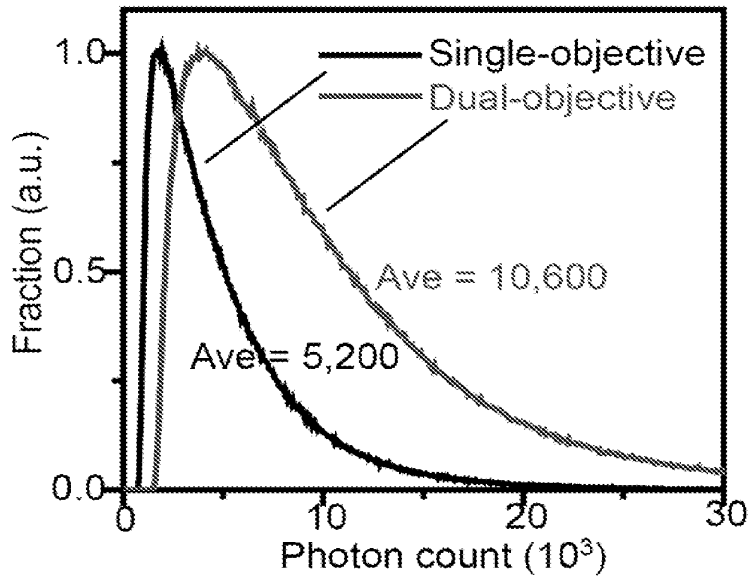

Single-objective
Dual-objective
Ave = 10,600
Ave = 5,200
Fraction (a.u.)
Photon count (10³)

Fig. 1D

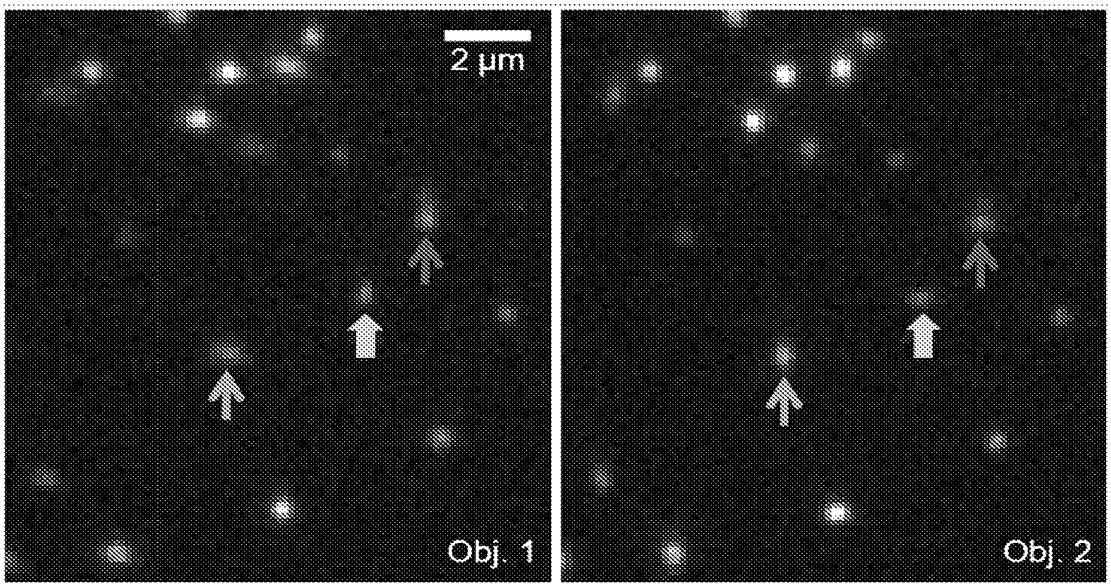


Fig. 1E

Fig. 2A

Fig. 2B

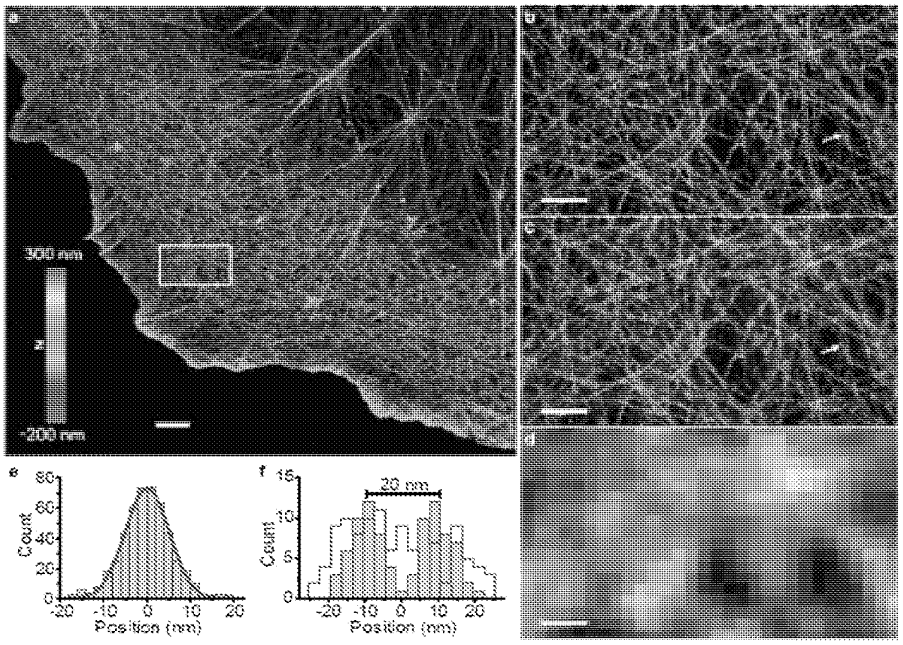


Fig. 2C

Fig. 2E

Fig. 2F

Fig. 2D

Fig. 3A

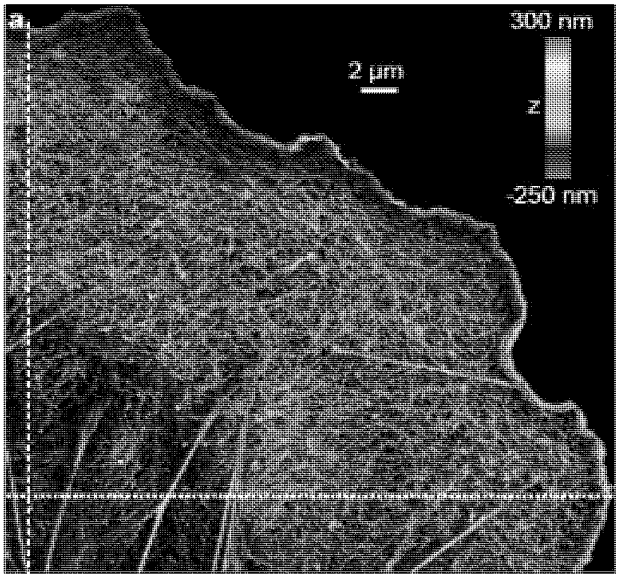


Fig. 3B

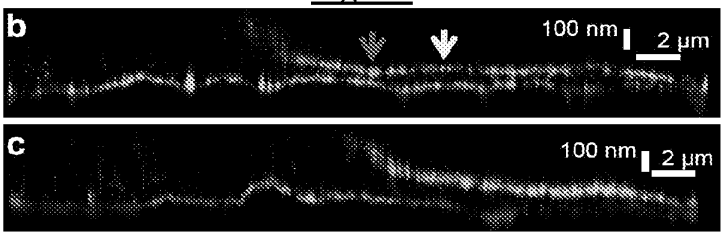


Fig. 3C

Fig. 3D

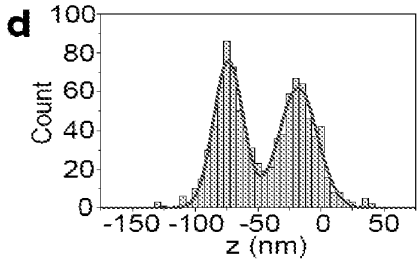
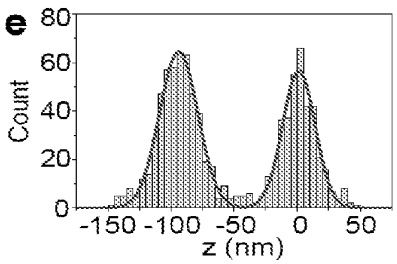


Fig. 3E



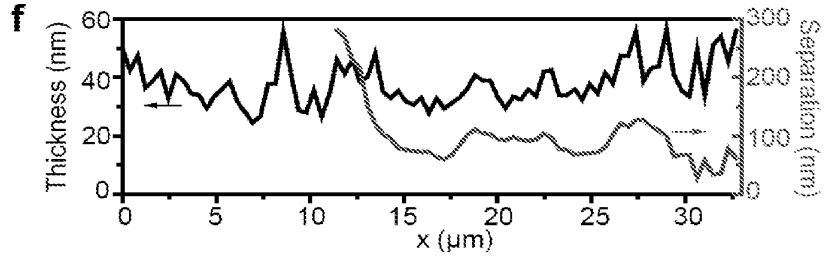


Fig. 3F

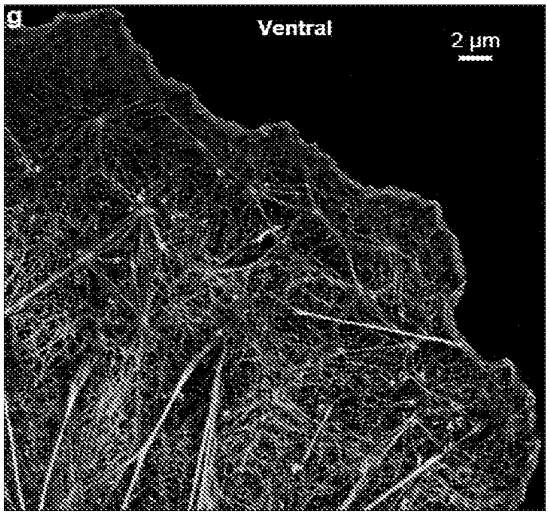


Fig. 3G

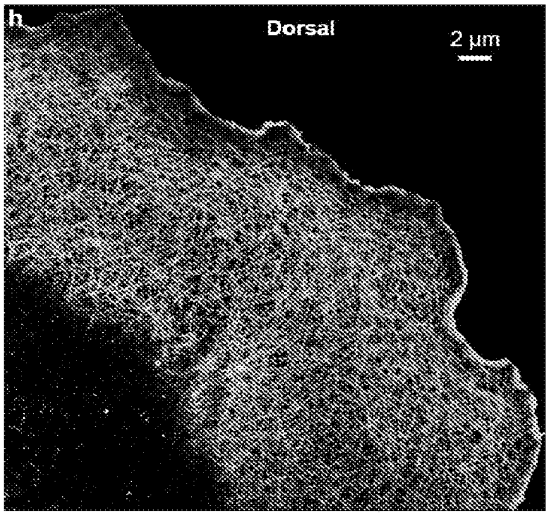


Fig. 3H

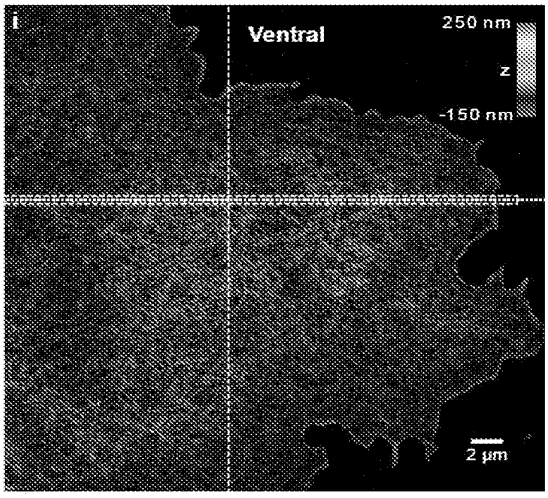


Fig. 3I

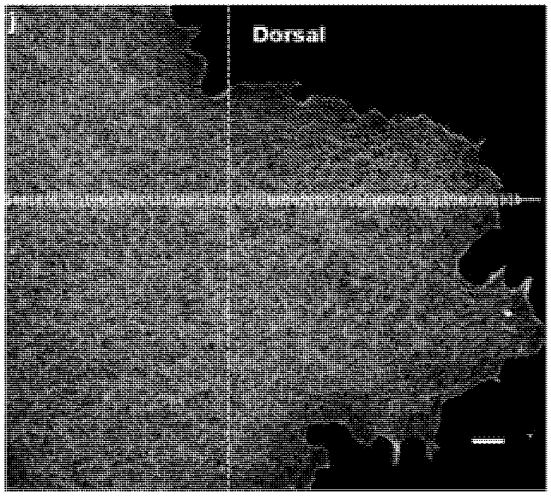


Fig. 3J

Fig. 3K

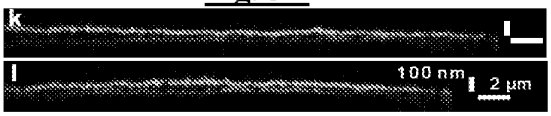


Fig. 3L

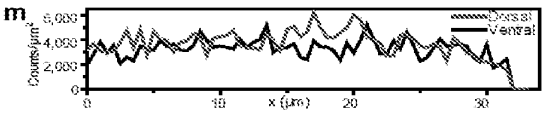


Fig. 3M

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 February 2016 (04.02.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/018960 A1

(51) International Patent Classification:
C40B 30/04 (2006.01) *C12Q 1/68* (2006.01)
C07H 21/02 (2006.01)

(21) International Application Number:
PCT/US2015/042556

(22) International Filing Date:
29 July 2015 (29.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/031,062 30 July 2014 (30.07.2014) US
62/050,636 15 September 2014 (15.09.2014) US
62/142,653 3 April 2015 (03.04.2015) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **ZHUANG, Xiaowei**; 5 Luongo Farm Lane, Lexington, MA 02421 (US). **CHEN, Kok-Hao**; 37 Jalan Lepan, Singapore 537961 (SG). **BOETTIGER, Alistair**; 1550 Cambridge Street, Apt. #2, Cambridge, MA 02139 (US). **MOFFITT, Jeffrey, R.**; 175 Beacon Street, Apt. #108, Somerville, MA 02143 (US). **WANG, Siyuan**; 28 Fernald Drive, Apt. #31, Cambridge, MA 02138 (US).

(74) Agent: **CHEN, Tani**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

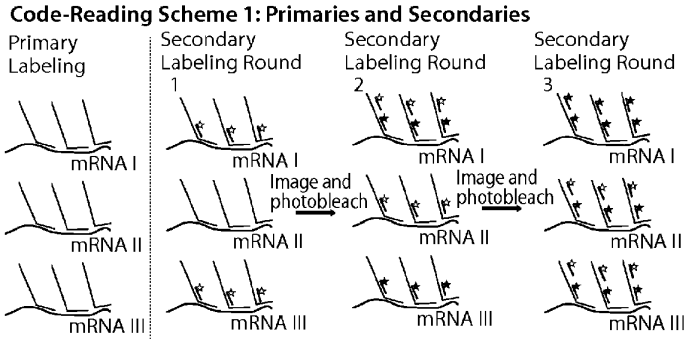


Fig. 1B

(57) Abstract: The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

WO 2016/018960 A1

SYSTEMS AND METHODS FOR DETERMINING NUCLEIC ACIDS

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/031,062, filed July 30, 2014, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; U.S. Provisional Patent Application Serial No. 62/142,653, filed April 3, 2015, entitled "Systems and Methods for Determining Nucleic Acids," by Zhuang, *et al.*; and U.S. Provisional Patent Application Serial No. 62/050,636, filed September 15, 2014, entitled "Probe Library Construction," by Zhuang, *et al.* Each of the above is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with government support under Grant No. GM096450 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined.

BACKGROUND

Single-molecule fluorescent in situ hybridization (smFISH) is a powerful method for detecting individual mRNA molecules in cells. The high detection efficiency and large dynamic range of this method provides exquisite detail into the expression state, spatial distribution within cells and intact tissues, and variation among cells of individual mRNAs. Such approaches have been essential to many recent insights into understanding gene regulation and expression. A fundamental limitation of smFISH, however, is its low throughput, typically only a few genes at a time. This low throughput is due to a lack of distinguishable probes with which to label cells and the cost of producing large amounts of labeled probe required for high efficiency staining. Thus, improvements in detecting mRNA molecules are needed.

SUMMARY

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

In one aspect, the present invention is generally directed to a composition. According to one set of embodiments, the composition comprises a plurality of nucleic acid probes, at least some of which comprise a first portion comprising a target sequence and a plurality of read sequences. In some cases, each comprises a first portion comprising a target sequence and a plurality of read sequences. In some embodiments, the plurality of read sequences are distributed on the plurality of nucleic acid probes so as to define an error-correcting code.

In another aspect, the present invention is generally directed to a method. In one set of embodiments, the method includes acts of exposing a sample to a plurality of nucleic acid probes; for each of the nucleic acid probes, determining binding of the nucleic acid probes within the sample; creating codewords based on the binding of the nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

The method, in another set of embodiments, includes acts of exposing a sample to a plurality of nucleic acid probes, wherein the nucleic acid probes comprise a first portion comprising a target sequence and a second portion comprising one or more read sequences, and wherein at least some of the plurality of nucleic acid probes comprises distinguishable nucleic acid probes formed from combinatorial combination of one or more read sequences taken from a plurality of read sequences; and for each of the nucleic acid probes, determining binding of the target sequences of the nucleic acid probes within the sample.

In yet another set of embodiments, the method includes acts of exposing a sample to a plurality of primary nucleic acid probes (also called encoding probes); exposing the plurality of primary nucleic acid probes to a sequence of secondary nucleic acid probes (also called readout probes) and determining fluorescence of each of the secondary nucleic acid probes within the sample; creating codewords based on fluorescence of the secondary nucleic acid probes; and for at least some of the codewords, matching the codeword to a valid codeword wherein, if no match is found, applying error correction to the codeword to form a valid codeword.

In one set of embodiments, the method includes acts of exposing a plurality of primary nucleic acid probes to a sample; and exposing the plurality of nucleic acid probes to a sequence of secondary nucleic acid probes and determining fluorescence of each of the secondary probes within the sample. In some embodiments, at least some of the plurality of secondary nucleic acid probes comprises distinguishable secondary nucleic acid probes formed from combinatorial combination of one or more read sequences (or readout probe sequences) taken from a plurality of read sequences (or readout probe sequences).

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In another set of embodiments, the method comprises acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a resolution better than 100 nm.

The method, in yet another set of embodiments, includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In certain embodiments, the method comprises acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords; and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

In addition, in one set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes corresponds to at least one common value of a position within the codewords.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions that is less than the number of targets, and wherein each codeword is associated with a single target, associating a plurality of distinguishable read sequences with the plurality of codewords such that each distinguishable read sequence represents a value of a position within the codewords, and forming a plurality of nucleic acid probes, each comprising a target sequence and one or more read sequences.

The method, in still another set of embodiments, includes acts of exposing a plurality of nucleic acid probes to a cell, exposing the plurality of nucleic acid probes to a sequence of secondary probes and determining fluorescence of each of the secondary probes within the cell, and based on the sequence of fluorescence of each of the secondary probes, determining nucleic acids within the cell.

In another set of embodiments, the method includes acts of associating a plurality of targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes each comprising a target sequence; and forming groups comprising the plurality of nucleic acid probes such that each group of nucleic acid probes correspond to at least one common value of a position within the codewords.

In yet another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity using a super-resolution imaging technique, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first secondary probe comprising a first signaling entity, determining the first signaling entity at a precision better than 500 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity at a precision better than 500 nm.

In still another set of embodiments, the method includes acts of exposing a cell to a plurality of nucleic acid probes, exposing the plurality of nucleic acid probes to a first

secondary probe comprising a first signaling entity, determining the first signaling entity at a resolution better than 100 nm, inactivating the first signaling entity, exposing the plurality of nucleic acid probes to a second secondary probe comprising a second signaling entity, and determining the second signaling entity using a super-resolution imaging technique.

In still another set of embodiments, the method includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords, wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code; associating unique read sequences with each possible value of each position in the codewords, wherein the read sequences are taken from a set of orthogonal sequences, which have limited homology with one another and with the nucleic acid species in a sample; forming a plurality of primary nucleic acid probes each comprising a target sequence that uniquely binds to a nucleic acid target and one or more read sequences; forming a plurality of secondary nucleic acid probes comprising a signaling entity and a sequence that is complementary to one of the read sequences; exposing a sample to the primary nucleic acid probes such that the nucleic acid probes hybridize to the nucleic acid targets in the sample; exposing the primary nucleic acid probes in the sample to a secondary nucleic acid probe such that the secondary nucleic acid probe hybridizes to the read sequence on at least some of the primary nucleic acid probes; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different secondary nucleic acid probe for at least some of the repetitions.

The method, according to yet another set of embodiments, includes acts of associating a plurality of nucleic acid targets with a plurality of target sequences and a plurality of codewords wherein the codewords comprise a number of positions and values for each position, and the codewords form an error-checking and/or error-correcting code space; forming a plurality of nucleic acid probes comprising a signaling entity and a target sequence that uniquely binds to one of the nucleic acid targets; grouping the nucleic acid probes into a plurality of probe pools, wherein each of the probe pools corresponds to a specific value of a unique position within the codewords; exposing a sample to one of the probe pools; imaging the sample; and repeating the exposing and imaging steps one or more times, using a different probe pool for at least some of the repetitions.

In another aspect, the present invention encompasses methods of making one or more of the embodiments described herein. In still another aspect, the present invention encompasses methods of using one or more of the embodiments described herein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF THE DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1C illustrates an encoding scheme for nucleic acid probes, in certain embodiments of the invention;

Figs. 2A-2G illustrates determining of mRNAs in a cell, in some embodiments of the invention;

Figs. 3A-3B illustrate the determination of nucleic acids, in accordance with various embodiments of the invention;

Figs. 4A-4B is a non-limiting example of multiple read sequences distributed in a population of different nucleic acid probes, in accordance with certain embodiments of the invention;

Figs. 5A-5E illustrate the determination of nucleic acids, in accordance with another embodiment of the invention;

Figs. 6A-6H illustrate simultaneous determination of multiple nucleic acid species in cells, in certain embodiments of the invention;

Figs. 7A-7E show expression noise of genes and co-variation of expression between different genes determined in accordance with some embodiments of the invention;

Figs. 8A-8E illustrate spatial distribution of RNAs in cells determined in accordance with one embodiment of the invention;

Figs. 9A-9C illustrate simultaneous determination of multiple nucleic acid species in cells, in another embodiment of the invention;

Figs. 10A-10B show expression between different genes determined in accordance with yet another embodiment of the invention;

Fig. 11 is a schematic description of combinatorial labeling, in accordance with another embodiment of the invention;

Fig. 12 shows schematic descriptions of Hamming distance, in another embodiment of the invention;

Fig. 13 illustrates the production of a library of probes, in still another embodiment of the invention;

Figs. 14A-14B illustrate fluorescent spot determinations, in another embodiment of the invention;

Figs. 15A-15B illustrate error correction facilitates RNA detection, in yet another embodiment of the invention;

Figs. 16A-16B show characterization of misidentification rates and calling rates, in one embodiment of the invention;

Figs. 17A-17D show characterization of misidentification rates and calling rates, in another embodiment of the invention;

Figs. 18A-18C show a comparison of experiments, in accordance with one embodiment of the invention; and

Figs. 19A-9D illustrate decoding and error assessment, in another embodiment of the invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO: 1 is:
GTTGGCGACGAAAGCACTGCGATTGGAACCGTCCCAAGCGTTGCGCTTAATGGA
TCATCAATTTTGTCTCACTACGACGGTCAATCGCGCTGCATACTTGCGTCGGTCG
GACAAACGAGG;

SEQ ID NO: 2 is CGCAACGCTTGGGACGGTTCCAATCGGATC;

SEQ ID NO: 3 is CGAATGCTCTGGCCTCGAACGAACGATAGC;

SEQ ID NO: 4 is ACAAATCCGACCAGATCGGACGATCATGGG;

SEQ ID NO: 5 is CAAGTATGCAGCGCGATTGACCGTCTCGTT;

SEQ ID NO: 6 is TGCGTCGTCTGGCTAGCACGGCACGCAAAT;

SEQ ID NO: 7 is AAGTCGTACGCCGATGCGCAGCAATTCACT;

SEQ ID NO: 8 is CGAAACATCGGCCACGGTCCCGTTGAACTT;

SEQ ID NO: 9 is ACGAATCCACCGTCCAGCGCGTCAAACAGA;

SEQ ID NO: 10 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC;

SEQ ID NO: 11 is GCATGAGTTGCCTGGCGTTGCGACGACTAA;

SEQ ID NO: 12 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC;

SEQ ID NO: 13 is GGCCAATGGCCCAGGTCCGTCACGCAATTT;

SEQ ID NO: 14 is TTGATCGAATCGGAGCGTAGCGGAATCTGC;

SEQ ID NO: 15 is CGCGCGGATCCGCTTGTCGGGAACGGATAC;

SEQ ID NO: 16 is GCCTCGATTACGACGGATGTAATTCGGCCG;

SEQ ID NO: 17 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT

SEQ ID NO: 18 is GTTGGTCGGCACTTGGGTGC;

SEQ ID NO: 19 is CGATGCGCCAATTCCGGTTC;

SEQ ID NO: 20 is CGCGGGCTATATGCGAACCG;

SEQ ID NO: 21 is TAATACGACTCACTATAGGGAAAGCCGGTTCATCCGGTGG;

SEQ ID NO: 22 is TAATACGACTCACTATAGGGTGATCATCGCTCGCGGGTTG;

SEQ ID NO: 23 is TAATACGACTCACTATAGGGCGTGGAGGGCATACAACGC;

SEQ ID NO: 24 is CGCAACGCTTGGGACGGTTCCAATCGGATC/3Cy5Sp/;

SEQ ID NO: 25 is CGAATGCTCTGGCCTCGAACGAACGATAGC/3Cy5Sp/;

SEQ ID NO: 26 is ACAAATCCGACCAGATCGGACGATCATGGG/3Cy5Sp/;

SEQ ID NO: 27 is CAAGTATGCAGCGCGATTGACCGTCTCGTT/3Cy5Sp/;

SEQ ID NO: 28 is GCGGGAAGCACGTGGATTAGGGCATCGACC/3Cy5Sp/;

SEQ ID NO: 29 is AAGTCGTACGCCGATGCGCAGCAATTCACT/3Cy5Sp/;

SEQ ID NO: 30 is CGAAACATCGGCCACGGTCCCGTTGAACTT/3Cy5Sp/;

SEQ ID NO: 31 is ACGAATCCACCGTCCAGCGCGTCAAACAGA/3Cy5Sp/;

SEQ ID NO: 32 is CGCGAAATCCCCGTAACGAGCGTCCCTTGC/3Cy5Sp/;

SEQ ID NO: 33 is GCATGAGTTGCCTGGCGTTGCGACGACTAA/3Cy5Sp/;

SEQ ID NO: 34 is CCGTCGTCTCCGGTCCACCGTTGCGCTTAC/3Cy5Sp/;

SEQ ID NO: 35 is GGCCAATGGCCCAGGTCCGTCACGCAATTT/3Cy5Sp/;

SEQ ID NO: 36 is TTGATCGAATCGGAGCGTAGCGGAATCTGC/3Cy5Sp/;

SEQ ID NO: 37 is CGCGCGGATCCGCTTGTCGGGAACGGATAC/3Cy5Sp/;

SEQ ID NO: 38 is GCCTCGATTACGACGGATGTAATTCGGCCG/3Cy5Sp/; and

SEQ ID NO: 39 is GCCCGTATTCCCGCTTGCGAGTAGGGCAAT/3Cy5Sp/

.

DETAILED DESCRIPTION

The present invention generally relates to systems and methods for imaging or determining nucleic acids, for instance, within cells. In some embodiments, the transcriptome of a cell may be determined. Certain embodiments are directed to determining nucleic acids, such as mRNA, within cells at relatively high resolutions. In some embodiments, a plurality of nucleic acid probes may be applied to a sample, and their binding within the sample determined, e.g., using fluorescence, to determine locations of the nucleic acid probes within the sample. In some embodiments, codewords may be based on the binding of the plurality of nucleic acid probes, and in some cases, the codewords may define an error-correcting code to reduce or prevent misidentification of the nucleic acids. In certain cases, a relatively large number of different targets may be identified using a relatively small number of labels, e.g., by using various combinatorial approaches.

Two example approaches are now discussed. It should be understood, however, that these are presented by way of explanation and not limitation; other aspects and embodiments are discussed in further detail herein. In one example method, primary probes (also called encoding probes) and secondary probes (also called readout probes) are used, where the primary probes encode "codewords" and bind to target nucleic acids in the sample, and the secondary probes are used to read out the codewords from the primary probes. In another example method, a plurality of different primary probes containing codewords are divided into as many separate pools as there are positions in the codewords, such that each primary probe pool corresponds to a certain value in a certain position of the codewords (e.g., a "one" in the first position as in "1001").

The first example is now described with respect to Fig. 3A. As will be discussed in more detail below, in other embodiments, other configurations may be used as well. In this first example, a series of nucleic acid probes are used to determine nucleic acids within a cell or other sample, e.g., qualitatively or quantitatively. For example, nucleic acids may be identified as being present or absent, and/or the numbers or concentrations of certain nucleic acids may be determined within the cell or other sample. In some cases, the positions of the probes within the cell or other sample may be determined at relatively high resolutions, and in some cases, at resolutions better than the wavelength of visible light.

This example is generally directed to spatially detecting nucleic acids within a cell or other sample, e.g., at relatively high resolutions. For example, the nucleic acids may be mRNAs, or other nucleic acids described herein. In one set of embodiments, the nucleic acids within the cell may be determined by delivering or applying nucleic acid probes to the

cell. In some cases, by using combinatorial approaches, a relatively large number of nucleic acids may be determined using a relatively small number of different labels on the nucleic acid probes. Thus, for example, a relatively small number of experiments may be used to determine a relatively large number of nucleic acids in a sample, e.g., due to simultaneous binding of the nucleic acid probes to different nucleic acids in the sample.

In one set of embodiments, a population of primary nucleic acid probes are applied to the cell (or other sample) that is able to bind nucleic acids suspected of being present within the cell. Afterwards, sequentially, secondary nucleic acid probes that can bind to or otherwise interact with some of the primary nucleic acids are added and determined, e.g., using imaging techniques such as fluorescence microscopy (e.g., conventional fluorescence microscopy), STORM (stochastic optical reconstruction microscopy) or other imaging techniques. After imaging, the secondary nucleic acid probes are inactivated or removed, and a different secondary nucleic acid probe is added to the sample. This may be repeated multiple times with multiple different secondary nucleic acid probes. The pattern of binding of the various secondary nucleic acid probes may be used to determine the primary nucleic acid probes at locations within the cell or other sample, which can be used to determine mRNA or other nucleic acids that are present.

For instance, as is shown in Fig. 3A, a population of nucleic acids 10 within a cell (represented here by nucleic acids 11, 12, and 13) may be exposed to a population of primary nucleic acid probes 20, including probes 21 and 22. The primary nucleic acid probes may contain, for instance, a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11). Probes 21 and 22 may contain the same or different targeting sequences, which may bind to or hybridize with the same or different nucleic acids. As an example, as is shown in Fig. 3A, probe 21 contains a first targeting sequence 25 which targets the probe to nucleic acid 11, while probe 22 contains a second targeting sequence 26, not identical to the first targeting sequence 25 and which targets the probe to nucleic acid 12. The target sequence may be substantially complementary to at least a portion of a target nucleic acid, and enough of the target sequence may be present such that specific binding of the nucleic acid probe to the target nucleic acid can occur.

Primary nucleic acid probes 20 may also contain one or more "read" sequences. Two such read sequences are used in this example, although in other embodiments, there may be one, three, four, or more read sequences present within a primary nucleic acid probe. The read sequences may all independently be the same or different. In addition, in one set of embodiments, different nucleic acid probes may use one or more common read sequences.

For example, more than one read sequence may be combinatorially combined on different nucleic acid probes, thereby producing a relatively large number of different nucleic acid probes that can be separately identified, even though only a relatively small number of read sequences are used. Thus, for example, in Fig. 3A, probe 21 contains read sequences 27 and 29, while probe 22 contains read sequences 27 and 28, where the two read sequences 27 are identical, and different from read sequences 28 and 29.

After primary nucleic acid probes 20 have been introduced to the sample and allowed to interact with nucleic acids 11, 12, and 13, one or more secondary nucleic acid probes 30 may be applied to the sample to determine the primary nucleic acid probes. The secondary nucleic acid probes may contain a recognition sequence able to recognize one of the read sequences present within the population of primary nucleic acid probes. For instance, the recognition sequence may be substantially complementary to at least a portion of the read sequence, such that the secondary nucleic acid probe is able to bind to or hybridize with corresponding primary nucleic acid probe. For instance, in this example, recognition sequence 35 is able to recognize read sequence 27. In addition, the secondary nucleic acid probes may contain one or more signaling entities 33. For example, a signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion. More than one secondary sequence may be used, e.g., sequentially. For example, as shown in this figure, the initial secondary probe 30 may be removed (e.g., as discussed below) and a new secondary probe 31 may be added, containing recognition sequence 36 able to recognize read sequence 28 and one or more signaling entities 33. This may also be repeated multiple times, e.g., to determine read sequence 29 or other read sequences that may be present.

The location of the secondary nucleic acid probes 30, 31, etc. may be determined by determining signaling entity 33. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity. In some embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques (e.g., resolutions better than the wavelength of visible light or the diffraction limit of light) may be used. Examples of super-resolution imaging techniques include STORM, or other techniques as discussed herein. In some cases, e.g., with certain super-resolution imaging techniques such as STORM, more than one image of the sample may be acquired.

More than one type of secondary nucleic acid probe may be applied to a cell or other sample. For example, a first secondary nucleic acid probe may be applied that can recognize

a first read sequence, then it or its attached signaling entity may be inactivated or removed, and a second secondary nucleic acid probe may be applied that can recognize a second read sequence. This process may be repeated multiple times, each with a different secondary nucleic acid probe, e.g., to determine the read sequences that were present in the various primary nucleic acid probes. Thus, primary nucleic acids within the sample can be determined on the basis of the binding pattern of secondary nucleic acid probes.

For example, a first location within the cell or other sample may exhibit binding of a first secondary probe and a third secondary probe, but not the binding of a second or a fourth secondary probe, while a second location may exhibit a different pattern of binding of various secondary probes. The primary nucleic acid probe that the secondary probes are able to bind to or hybridize with may be determined by considering the pattern of binding of various secondary probes. For instance, referring to Fig. 3A, if a first secondary probe is able to determine read sequence 27, a second secondary probe is able to determine read sequence 28, and a third secondary probe is able to determine read sequence 29, then primary nucleic acid 25 may be determined through the binding of the first and third secondary probes (but not the second secondary probe), while primary nucleic acid 26 may be determined through the binding of the first and second secondary probes (but not the third secondary probe). Similarly, if it is known that first probe 21 contains target sequence 25 while second probe 22 contains target sequence 26, then nucleic acids 11 and 12 may also be determined within the sample, e.g., spatially, based on the binding pattern of the various secondary nucleic acid probes. In addition, it should be noted that due to the presence of more than one read sequence on the primary nucleic acid probes, even though first probe 21 and second probe 22 contains a common read sequence (read sequence 27), these probes may be distinguished in the sample due to the different binding patterns of the various secondary nucleic acid probes.

In certain embodiments, this pattern of binding or hybridization of the secondary nucleic acid probes may be converted into a "codeword." In this example, for instance, the codewords are "101" and "110" for first probe 21 and second probe 22, respectively, where a value of 1 represents binding and a value of 0 represents no binding. The codewords may also have longer lengths in other embodiments; only three probes are shown here for clarity purposes only. A codeword can be directly related to a specific target nucleic acid sequence of the primary nucleic acid probe. Accordingly, different primary nucleic acid probes may match certain codewords, which can then be used to identify the different targets of the primary nucleic acid probes based on the binding patterns of the secondary probes, even if in some cases, there is overlap in the read sequences of different secondary probes, e.g., as was

shown in Fig. 3A. However, if no binding is evident (e.g., for nucleic acid 13), then the codeword would be "000" in this example.

The values in each codeword can also be assigned in different fashions in some embodiments. For example, a value of 0 could represent binding while a value of 1 represents no binding. Similarly, a value of 1 could represent binding of a secondary nucleic acid probe with one type of signaling entity while a value of 0 could represent binding of a secondary nucleic acid probe with another type of distinguishable signaling entity. These signaling entities could be distinguished, for example, via different colors of fluorescence. In some cases, values in codewords need not be confined to 0 and 1. The values could also be drawn from larger alphabets, such as ternary (e.g., 0, 1, and 2) or quaternary (e.g., 0, 1, 2, and 3) systems. Each different value could, for example, be represented by a different distinguishable signaling entity, including (in some cases) one value that may be represented by the absence of signal.

The codewords for each target may be assigned sequentially, or may be assigned at random. For instance, referring to Fig. 3A, a first nucleic acid target may be assigned to 101, while a second nucleic acid target may be assigned to 110. In addition, in some embodiments, the codewords may be assigned using an error-detection system or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system (or a SECDED system, i.e., single error correction, double error detection). Generally speaking, such systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been. For example, a codeword such as 001 may be detected as invalid and corrected using such a system to 101, e.g., if 001 is not previously assigned to a different target sequence. A variety of different error-correcting codes can be used, many of which have previously been developed for use within the computer industry; however, such error-correcting systems have not typically been used within biological systems. Additional examples of such error-correcting codes are discussed in more detail below.

It should also be understood that all possible codewords in a code need not be used in some cases. For example, in some embodiments, codewords that are not used can serve as negative controls. Similarly, in some embodiments, some codewords can be left out because they are more prone to errors in measurement than other codewords. For example, in some implementations, reading a codeword with more values of '1' might be more error-prone that reading a codeword with fewer values of '1.'

It should be understood that the above description is an example of one embodiment of the invention, and that primary and secondary nucleic acid probes are not necessary in all embodiments. For example, in some embodiments, a series of nucleic acid probes containing signaling entities are used to determine nucleic acids within a cell or other sample, without necessarily requiring secondary probes.

For example, turning now to Fig. 3B, nucleic acids 11, 12, and 13 are exposed to different rounds of probes 21, 22, 23, 24, etc. in this example. These probes may each contain a target sequence that can recognize a nucleic acid (e.g., a sequence within nucleic acid 11 or 12). These probes may each target the same nucleic acid, but different regions of the nucleic acid. In addition, some or all of the probes may contain one or more signaling entities, e.g., signaling entity 29 on probe 21. For example, the signaling entity may be a fluorescent entity attached to the probe, or a certain sequence of nucleic acids that can be determined in some fashion.

The first round of probes (e.g. probe 21 and probe 22) may be applied to the cell or other sample. Probe 21 may be allowed to bind to nucleic acid 11 via target sequence 25. Such binding can be determined by determining signaling entity 29. For example, if the signaling entity is fluorescent, then fluorescence microscopy can be used to determine the signaling entity, e.g., spatially within the cell or other sample. In some but not all embodiments, imaging of a sample to determine the signaling entity may be used at relatively high resolutions, and in some cases, super-resolution imaging techniques may be used. Other, different probes may be present as well; for instance, probe 22 containing target sequence 26 may bind to nucleic acid 12, and be determined via signaling entity 29 within probe 22. These may occur, e.g., sequentially or simultaneously. Optionally, probes 21 and 22 may also be removed or inactivated, e.g., between application of different rounds of probes.

Next, a second round of probes (e.g., probe 23) is applied to the sample. In this example, probe 23 is able to bind to nucleic acid 11 via a targeting region, although there is no probe in the second round that is able to bind to nucleic acid 12. Binding of the probes is allowed to occur as discussed above, and determination of binding may occur via signaling entities. These signaling entities may be the same or different as from the first round of probes. This process may be repeated any number of times with different probes. For example, as is shown in Fig. 3B, round 2 contains probes able to bind to nucleic acid 11, while round 3 contains probes able to bind to nucleic acid 12.

In certain embodiments, each round of binding or hybridization of nucleic acid probes may be converted into a "codeword." In this example, using probes 21, 22, 23, and 24, the codewords 101 or 110 could be formed, where 1 represents binding and 0 represents no binding and the first position corresponds to the binding of probes 21 or 22 while the second position corresponds to the binding of probes 22, and the third position corresponds to the binding of probe 24. A codeword of 000 would represent no binding, e.g., as shown with nucleic acid 13 in this example. A codeword can be directly related to a specific target nucleic acid sequence of the nucleic acid probes, by designing appropriate nucleic acid probes. Thus, for example, 110 may correspond to a first target nucleic acid 12 (e.g., the first and second round of nucleic acid probes containing probes able to target nucleic acid 11, and these probes may target the same or different regions of nucleic acid 11) while 101 may correspond to a second target nucleic acid (e.g., the first and third round of nucleic acid probes containing probes able to target nucleic acid 12, and these probes may target the same or different regions of nucleic acid 12). In addition, it should be noted that each round of probes may contain the same, or different signaling entities as other probes in the same round, and/or other probes in different rounds. For instance, in one set of embodiments, only one signaling entity is used in all of the rounds of probes.

Similar to the above, the codewords for each target may be assigned sequentially, or may be assigned at random. The codewords may be assigned within a code space in some embodiments using an error-detection or an error-correcting system, such as a Hamming system, a Golay code, or an extended Hamming system or a SECDED system (single error correction, double error detection). Generally speaking, such error-correction systems can be used to identify where errors have occurred, and in some cases, such systems can also be used to correct the errors and determine what the correct codeword should have been.

Similar to the above, the values at each position in the codeword can be arbitrarily assigned in certain embodiments to binding or non-binding of probes that contain more than one distinguishable signaling entity.

In some cases, the nucleic acid probes may be formed into "pools" or groups of nucleic acids that share a common feature. For example, probes to all targets with codewords that contain a 1 in the first position, e.g. 110 and 101 but not 011, may comprise one pool while probes to all targets that contain a 1 in the second position, e.g. 110 and 011 but not 101, may comprise another pool. See also Fig. 1C. In some cases, a nucleic acid probe may be a member of more than one group or pool. Members of a nucleic acid pool may also contain features in addition to target sequences, read sequences, and or signaling entities that

allow them to be distinguished from other groups. These features may be short nucleic acid sequences that are used for the amplification, production, or separation of these sequences. The nucleic acid probes of each group may be applied to a sample, e.g., sequentially, as discussed herein.

Thus, in some aspects, the present invention is generally directed to systems and methods for determining nucleic acids within a cell or other sample. The sample may include a cell culture, a suspension of cells, a biological tissue, a biopsy, an organism, or the like. The sample may also be cell-free but nevertheless contain nucleic acids. If the sample contains a cell, the cell may be a human cell, or any other suitable cell, e.g., a mammalian cell, a fish cell, an insect cell, a plant cell, or the like. More than one cell may be present in some cases.

The nucleic acids to be determined may be, for example, DNA, RNA, or other nucleic acids that are present within a cell (or other sample). The nucleic acids may be endogenous to the cell, or added to the cell. For instance, the nucleic acid may be viral, or artificially created. In some cases, the nucleic acid to be determined may be expressed by the cell. The nucleic acid is RNA in some embodiments. The RNA may be coding and/or non-coding RNA. Non-limiting examples of RNA that may be studied within the cell include mRNA, siRNA, rRNA, miRNA, tRNA, lncRNA, snoRNAs, snRNAs, exRNAs, piRNAs, or the like.

In some cases, a significant portion of the nucleic acid within the cell may be studied. For instance, in some cases, enough of the RNA present within a cell may be determined so as to produce a partial or complete transcriptome of the cell. In some cases, at least 4 types of mRNAs are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 types of mRNAs may be determined within a cell.

In some cases, the transcriptome of a cell may be determined. It should be understood that the transriptome generally encompasses all RNA molecules produced within a cell, not just mRNA. Thus, for instance, the transcriptome may also include rRNA, tRNA, siRNA, etc. In some embodiments, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%,

at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% of the transcriptome of a cell may be determined.

The determination of one or more nucleic acids within the cell or other sample may be qualitative and/or quantitative. In addition, the determination may also be spatial, e.g., the position of the nucleic acid within the cell or other sample may be determined in two or three dimensions. In some embodiments, the positions, number, and/or concentrations of nucleic acids within the cell (or other sample) may be determined.

In some cases, a significant portion of the genome of a cell may be determined. The determined genomic segments may be continuous or interspersed on the genome. For example, in some cases, at least 4 genomic segments are determined within a cell, and in some cases, at least 3, at least 4, at least 7, at least 8, at least 12, at least 14, at least 15, at least 16, at least 22, at least 30, at least 31, at least 32, at least 50, at least 63, at least 64, at least 72, at least 75, at least 100, at least 127, at least 128, at least 140, at least 255, at least 256, at least 500, at least 1,000, at least 1,500, at least 2,000, at least 2,500, at least 3,000, at least 4,000, at least 5,000, at least 7,500, at least 10,000, at least 12,000, at least 15,000, at least 20,000, at least 25,000, at least 30,000, at least 40,000, at least 50,000, at least 75,000, or at least 100,000 genomic segments may be determined within a cell.

In some cases, the entire genome of a cell may be determined. It should be understood that the genome generally encompasses all DNA molecules produced within a cell, not just chromosome DNA. Thus, for instance, the genome may also include, in some cases, mitochondria DNA, chloroplast DNA, plasmid DNA, etc. In some embodiments, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or 100% of the genome of a cell may be determined.

As discussed herein, a variety of nucleic acid probes may be used to determine one or more nucleic acids within a cell or other sample. The probes may comprise nucleic acids (or entities that can hybridize to a nucleic acid, e.g., specifically) such as DNA, RNA, LNA (locked nucleic acids), PNA (peptide nucleic acids), or combinations thereof. In some cases, additional components may also be present within the nucleic acid probes, e.g., as discussed below. Any suitable method may be used to introduce nucleic acid probes into a cell.

For example, in some embodiments, the cell is fixed prior to introducing the nucleic acid probes, e.g., to preserve the positions of the nucleic acids within the cell. Techniques for fixing cells are known to those of ordinary skill in the art. As non-limiting examples, a cell

may be fixed using chemicals such as formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, acetic acid, or the like. In one embodiment, a cell may be fixed using Hepes-glutamic acid buffer-mediated organic solvent (HOPE).

The nucleic acid probes may be introduced into the cell (or other sample) using any suitable method. In some cases, the cell may be sufficiently permeabilized such that the nucleic acid probes may be introduced into the cell by flowing a fluid containing the nucleic acid probes around the cells. In some cases, the cells may be sufficiently permeabilized as part of a fixation process; in other embodiments, cells may be permeabilized by exposure to certain chemicals such as ethanol, methanol, Triton, or the like. In addition, in some embodiments, techniques such as electroporation or microinjection may be used to introduce nucleic acid probes into a cell or other sample.

Certain aspects of the present invention are generally directed to nucleic acid probes that are introduced into a cell (or other sample). The probes may comprise any of a variety of entities that can hybridize to a nucleic acid, typically by Watson-Crick base pairing, such as DNA, RNA, LNA, PNA, etc., depending on the application. The nucleic acid probe typically contains a target sequence that is able to bind to at least a portion of a target nucleic acid, in some cases specifically. When introduced into a cell or other system, the target system may be able to bind to a specific target nucleic acid (e.g., an mRNA, or other nucleic acids as discussed herein). In some cases, the nucleic acid probes may be determined using signaling entities (e.g., as discussed below), and/or by using secondary nucleic acid probes able to bind to the nucleic acid probes (i.e., to primary nucleic acid probes). The determination of such nucleic acid probes is discussed in detail below.

In some cases, more than one type of (primary) nucleic acid probe may be applied to a sample, e.g., simultaneously. For example, there may be at least 2, at least 5, at least 10, at least 25, at least 50, at least 75, at least 100, at least 300, at least 1,000, at least 3,000, at least 10,000, or at least 30,000 distinguishable nucleic acid probes that are applied to a sample, e.g., simultaneously or sequentially.

The target sequence may be positioned anywhere within the nucleic acid probe (or primary nucleic acid probe or encoding nucleic acid probe). The target sequence may contain a region that is substantially complementary to a portion of a target nucleic acid. In some cases, the portions may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary. In some cases, the target sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40,

at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the target sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the target sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc. Typically, complementarity is determined on the basis of Watson-Crick nucleotide base pairing.

The target sequence of a (primary) nucleic acid probe may be determined with reference to a target nucleic acid suspected of being present within a cell or other sample. For example, a target nucleic acid to a protein may be determined using the protein's sequence, by determining the nucleic acids that are expressed to form the protein. In some cases, only a portion of the nucleic acids encoding the protein are used, e.g., having the lengths as discussed above. In addition, in some cases, more than one target sequence that can be used to identify a particular target may be used. For instance, multiple probes can be used, sequentially and/or simultaneously, that can bind to or hybridize to different regions of the same target. Hybridization typically refers to an annealing process by which complementary single-stranded nucleic acids associate through Watson-Crick nucleotide base pairing (e.g., hydrogen bonding, guanine-cytosine and adenine-thymine) to form double-stranded nucleic acid.

In some embodiments, a nucleic acid probe, such as a primary nucleic acid probe, may also comprise one or more "read" sequences. However, it should be understood that read sequences are not necessary in all cases. In some embodiments, the nucleic acid probe may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or more, 20 or more, 32 or more, 40 or more, 50 or more, 64 or more, 75 or more, 100 or more, 128 or more read sequences. The read sequences may be positioned anywhere within the nucleic acid probe. If more than one read sequence is present, the read sequences may be positioned next to each other, and/or interspersed with other sequences.

The read sequences, if present, may be of any length. If more than one read sequence is used, the read sequences may independently have the same or different lengths. For

instance, the read sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 50, at least 60, at least 65, at least 75, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, or at least 450 nucleotides in length. In some cases, the read sequence may be no more than 500, no more than 450, no more than 400, no more than 350, no more than 300, no more than 250, no more than 200, no more than 175, no more than 150, no more than 125, no more than 100, be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the read sequence may have a length of between 10 and 30 nucleotides, between 20 and 40 nucleotides, between 5 and 50 nucleotides, between 10 and 200 nucleotides, or between 25 and 35 nucleotides, between 10 and 300 nucleotides, etc.

The read sequence may be arbitrary or random in some embodiments. In certain cases, the read sequences are chosen so as to reduce or minimize homology with other components of the cell or other sample, e.g., such that the read sequences do not themselves bind to or hybridize with other nucleic acids suspected of being within the cell or other sample. In some cases, the homology may be less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or less than 1%. In some cases, there may be a homology of less than 20 basepairs, less than 18 basepairs, less than 15 basepairs, less than 14 basepairs, less than 13 basepairs, less than 12 basepairs, less than 11 basepairs, or less than 10 basepairs. In some cases, the basepairs are sequential.

In one set of embodiments, a population of nucleic acid probes may contain a certain number of read sequences, which may be less than the number of targets of the nucleic acid probes in some cases. Those of ordinary skill in the art will be aware that if there is one signaling entity and n read sequences, then in general 2^n-1 different nucleic acid targets may be uniquely identified. However, not all possible combinations need be used. For instance, a population of nucleic acid probes may target 12 different nucleic acid sequences, yet contain no more than 8 read sequences. As another example, a population of nucleic acids may target 140 different nucleic acid species, yet contain no more than 16 read sequences. Different nucleic acid sequence targets may be separately identified by using different combinations of read sequences within each probe. For instance, each probe may contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, etc. or more read sequences. In some cases, a population of nucleic acid probes may each contain the same number of read sequences, although in other cases, there may be different numbers of read sequences present on the various probes.

As a non-limiting example, a first nucleic acid probe may contain a first target sequence, a first read sequence, and a second read sequence, while a second, different nucleic acid probe may contain a second target sequence, the same first read sequence, but a third read sequence instead of the second read sequence. Such probes may thereby be distinguished by determining the various read sequences present or associated with a given probe or location, as discussed herein.

In addition, the nucleic acid probes (and their corresponding, complimentary sites on the encoding probes), in certain embodiments, may be made using only 2 or only 3 of the 4 bases, such as leaving out all the "G"s or leaving out all of the "C"s within the probe. Sequences lacking either "G"s or "C"s may form very little secondary structure in certain embodiments, and can contribute to more uniform, faster hybridization.

In some embodiments, the nucleic acid probe may contain a signaling entity. It should be understood that signaling entities are not required in all cases, however; for instance, the nucleic acid probe may be determined using secondary nucleic acid probes in some embodiments, as is discussed in additional detail below. Examples of signaling entities that can be used are also discussed in more detail below.

Other components may also be present within a nucleic acid probe as well. For example, in one set of embodiments, one or more primer sequences may be present, e.g., to allow for enzymatic amplification of probes. Those of ordinary skill in the art will be aware of primer sequences suitable for applications such as amplification (e.g., using PCR or other suitable techniques). Many such primer sequences are available commercially. Other examples of sequences that may be present within a primary nucleic acid probe include, but are not limited to promoter sequences, operons, identification sequences, nonsense sequences, or the like.

Typically, a primer is a single-stranded or partially double-stranded nucleic acid (e.g., DNA) that serves as a starting point for nucleic acid synthesis, allowing polymerase enzymes such as nucleic acid polymerase to extend the primer and replicate the complementary strand. A primer is (e.g., is designed to be) complementary to and to hybridize to a target nucleic acid. In some embodiments, a primer is a synthetic primer. In some embodiments, a primer is a non-naturally-occurring primer. A primer typically has a length of 10 to 50 nucleotides. For example, a primer may have a length of 10 to 40, 10 to 30, 10 to 20, 25 to 50, 15 to 40, 15 to 30, 20 to 50, 20 to 40, or 20 to 30 nucleotides. In some embodiments, a primer has a length of 18 to 24 nucleotides.

In addition, the components of the nucleic acid probe may be arranged in any suitable order. For instance, in one embodiment, the components may be arranged in a nucleic acid probe as: primer—read sequences—targeting sequence—read sequences—reverse primer. The "read sequences" in this structure may each contain any number (including 0) of read sequences, so long as at least one read sequence is present in the probe. Non-limiting example structures include primer—targeting sequence—read sequences—reverse primer, primer—read sequences—targeting sequence—reverse primer, targeting sequence—primer—targeting sequence—read sequences—reverse primer, targeting sequence—primer—read sequences—targeting sequence—reverse primer, primer—target sequence—read sequences—targeting sequence—reverse primer, targeting sequence—primer—read sequence—reverse primer, targeting sequence—read sequence—primer, read sequence—targeting sequence—primer, read sequence—primer—targeting sequence—reverse primer, etc. In addition, the reverse primer is optional in some embodiments, including in all of the above-described examples.

After introduction of the nucleic acid probes into a cell or other sample, the nucleic acid probes may be directly determined by determining signaling entities (if present), and/or the nucleic acid probes may be determined by using one or more secondary nucleic acid probes, in accordance with certain aspects of the invention. As mentioned, in some cases, the determination may be spatial, e.g., in two or three dimensions. In addition, in some cases, the determination may be quantitative, e.g., the amount or concentration of a primary nucleic acid probe (and of a target nucleic acid) may be determined. Additionally, the secondary probes may comprise any of a variety of entities able to hybridize a nucleic acid, e.g., DNA, RNA, LNA, and/or PNA, etc., depending on the application. Signaling entities are discussed in more detail below.

A secondary nucleic acid probe may contain a recognition sequence able to bind to or hybridize with a read sequence of a primary nucleic acid probe. In some cases, the binding is specific, or the binding may be such that a recognition sequence preferentially binds to or hybridizes with only one of the read sequences that are present. The secondary nucleic acid probe may also contain one or more signaling entities. If more than one secondary nucleic acid probe is used, the signaling entities may be the same or different.

The recognition sequences may be of any length, and multiple recognition sequences may be of the same or different lengths. If more than one recognition sequence is used, the recognition sequences may independently have the same or different lengths. For instance, the recognition sequence may be at least 5, at least 10, at least 15, at least 20, at least 25, at

least 30, at least 35, at least 40, or at least 50 nucleotides in length. In some cases, the recognition sequence may be no more than 75, no more than 60, no more than 65, no more than 60, no more than 55, no more than 50, no more than 45, no more than 40, no more than 35, no more than 30, no more than 20, or no more than 10 nucleotides in length. Combinations of any of these are also possible, e.g., the recognition sequence may have a length of between 10 and 30, between 20 and 40, or between 25 and 35 nucleotides, etc. In one embodiment, the recognition sequence is of the same length as the read sequence. In addition, in some cases, the recognition sequence may be at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% complementary to a read sequence of the primary nucleic acid probe.

As mentioned, in some cases, the secondary nucleic acid probe may comprise one or more signaling entities. Examples of signaling entities are discussed in more detail below.

As discussed, in certain aspects of the invention, nucleic acid probes are used that contain various "read sequences." For example, a population of primary nucleic acid probes may contain certain "read sequences" which can bind certain of the secondary nucleic acid probes, and the locations of the primary nucleic acid probes are determined within the sample using secondary nucleic acid probes, e.g., which comprise a signaling entity. As mentioned, in some cases, a population of read sequences may be combined in various combinations to produce different nucleic acid probes, e.g., such that a relatively small number of read sequences may be used to produce a relatively large number of different nucleic acid probes.

Thus, in some cases, a population of primary nucleic acid probes (or other nucleic acid probes) may each contain a certain number of read sequences, some of which are shared between different primary nucleic acid probes such that the total population of primary nucleic acid probes may contain a certain number of read sequences. A population of nucleic acid probes may have any suitable number of read sequences. For example, a population of primary nucleic acid probes may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 etc. read sequences. More than 20 are also possible in some embodiments. In addition, in some cases, a population of nucleic acid probes may, in total, have 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 20 or more, 24 or more, 32 or more, 40 or more, 50 or more, 60 or more, 64 or more, 100 or more, 128 or more, etc. of possible read sequences present, although some or all of the probes may each contain more than one read sequence, as discussed herein. In addition, in some embodiments,

the population of nucleic acid probes may have no more than 100, no more than 80, no more than 64, no more than 60, no more than 50, no more than 40, no more than 32, no more than 24, no more than 20, no more than 16, no more than 15, no more than 14, no more than 13, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 6, no more than 5, no more than 4, no more than 3, or no more than two read sequences present. Combinations of any of these are also possible, e.g., a population of nucleic acid probes may comprise between 10 and 15 read sequences in total.

As a non-limiting example of an approach to combinatorially producing a relatively large number of nucleic acid probes from a relatively small number of read sequences, in a population of 6 different types of nucleic acid probes, each comprising one or more read sequences, the total number of read sequences within the population may be no greater than 4. It should be understood that although 4 read sequences are used in this example for ease of explanation, in other embodiments, larger numbers of nucleic acid probes may be realized, for example, using 5, 8, 10, 16, 32, etc. or more read sequences, or any other suitable number of read sequences described herein, depending on the application. Referring now to Fig. 4A, if each of the primary nucleic acid probes contains two different read sequences, then by using 4 such read sequences (A, B, C, and D), up to 6 probes may be separately identified. It should be noted that in this example, the ordering of read sequences on a nucleic acid probe is not essential, i.e., "AB" and "BA" may be treated as being synonymous (although in other embodiments, the ordering of read sequences may be essential and "AB" and "BA" may not necessarily be synonymous). Similarly, if 5 read sequences are used (A, B, C, D, and E) in the population of primary nucleic acid probes, up to 10 probes may be separately identified, as is shown in Fig. 4B. For example, one of ordinary skill in the art would understand that, for k read sequences in a population with n read sequences on each probe, up to $\binom{n}{k}$ different probes may be produced, assuming that the ordering of read sequences is not essential; because not all of the probes need to have the same number of read sequences and not all combinations of read sequences need to be used in every embodiment, either more or less than this number of different probes may also be used in certain embodiments. In addition, it should also be understood that the number of read sequences on each probe need not be identical in some embodiments. For instance example, some probes may contain 2 read sequences while other probes may contain 3 read sequences.

In some aspects, the read sequences and/or the pattern of binding of nucleic acid probes within a sample may be used to define an error-detecting and/or an error-correcting

code, for example, to reduce or prevent misidentification or errors of the nucleic acids, e.g., as was discussed with reference to Fig. 3. Thus, for example, if binding is indicated (e.g., as determined using a signaling entity), then the location may be identified with a "1"; conversely, if no binding is indicated, then the location may be identified with a "0" (or vice versa, in some cases). Multiple rounds of binding determinations, e.g., using different nucleic acid probes, can then be used to create a "codeword," e.g., for that spatial location. In some embodiments, the codeword may be subjected to error detection and/or correction. For instance, the codewords may be organized such that, if no match is found for a given set of read sequences or binding pattern of nucleic acid probes, then the match may be identified as an error, and optionally, error correction may be applied sequences to determine the correct target for the nucleic acid probes. In some cases, the codewords may have fewer "letters" or positions that the total number of nucleic acids encoded by the codewords, e.g. where each codeword encodes a different nucleic acid.

Such error-detecting and/or the error-correction code may take a variety of forms. A variety of such codes have previously been developed in other contexts such as the telecommunications industry, such as Golay codes or Hamming codes. In one set of embodiments, the read sequences or binding patterns of the nucleic acid probes are assigned such that not every possible combination is assigned.

For example, if 4 read sequences are possible and a primary nucleic acid probe contains 2 read sequences, then up to 6 primary nucleic acid probes could be identified; but the number of primary nucleic acid probes used may be less than 6. Similarly, for k read sequences in a population with n read sequences on each primary nucleic acid probe, $\binom{n}{k}$ different probes may be produced, but the number of primary nucleic acid probes that are used may be any number more or less than $\binom{n}{k}$. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

As another example, if multiple rounds of nucleic acid probes are used, the number of rounds may be arbitrarily chosen. If in each round, each target can give two possible outcomes, such as being detected or not being detected, up to 2^n different targets may be possible for n rounds of probes, but the number of nucleic acid targets that are actually used may be any number less than 2^n. For example, if in each round, each target can give more than two possible outcomes, such as being detected in different color channels, more than 2^n (e.g. 3^n, 4^n ...) different targets may be possible for n rounds of probes. In some cases, the number of nucleic acid targets that are actually used may be any number less than this

number. In addition, these may be randomly assigned, or assigned in specific ways to increase the ability to detect and/or correct errors.

For example, in one set of embodiments, the codewords or nucleic acid probes may be assigned within a code space such that the assignments are separated by a Hamming distance, which measures the number of incorrect "reads" in a given pattern that cause the nucleic acid probe to be misinterpreted as a different valid nucleic acid probe. In certain cases, the Hamming distance may be at least 2, at least 3, at least 4, at least 5, at least 6, or the like. In addition, in one set of embodiments, the assignments may be formed as a Hamming code, for instance, a Hamming(7, 4) code, a Hamming(15, 11) code, a Hamming(31, 26) code, a Hamming(63, 57) code, a Hamming(127, 120) code, etc. In another set of embodiments, the assignments may form a SECDED code, e.g., a SECDED(8,4) code, a SECDED(16,4) code, a SCEDED(16, 11) code, a SCEDED(22, 16) code, a SCEDED(39, 32) code, a SCEDED(72, 64) code, etc. In yet another set of embodiments, the assignments may form an extended binary Golay code, a perfect binary Golay code, or a ternary Golay code. In another set of embodiments, the assignments may represent a subset of the possible values taken from any of the codes described above.

For example, a code with the same error correcting properties of the SECDED code may be formed by using only binary words that contain a fixed number of '1' bits, such as 4, to encode the targets. In another set of embodiments, the assignments may represent a subset of the possible values taken from codes described above for the purpose of addressing asymmetric readout errors. For example, in some cases, a code in which the number of '1' bits may be fixed for all used binary words may eliminate the biased measurement of words with different numbers of '1's when the rate at which '0' bits are measured as '1's or '1' bits are measured as '0's are different.

Accordingly, in some embodiments, once the codeword is determined (e.g., as discussed herein), the codeword may be compared to the known nucleic acid codewords. If a match is found, then the nucleic acid target can be identified or determined. If no match is found, then an error in the reading of the codeword may be identified. In some cases, error correction can also be applied to determine the correct codeword, and thus resulting in the correct identity of the nucleic acid target. In some cases, the codewords may be selected such that, assuming that there is only one error present, only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid target is possible. In some cases, this may also be generalized to larger codeword spacings or Hamming distances; for instance, the codewords may be selected such that if two, three, or four errors are present (or

more in some cases), only one possible correct codeword is available, and thus, only one correct identity of the nucleic acid targets is possible.

The error-correcting code may be a binary error-correcting code, or it may be based on other numbering systems, e.g., ternary or quaternary error-correcting codes. For instance, in one set of embodiments, more than one type of signaling entity may be used and assigned to different numbers within the error-correcting code. Thus, as a non-limiting example, a first signaling entity (or more than one signaling entity, in some cases) may be assigned as "1" and a second signaling entity (or more than one signaling entity, in some cases) may be assigned as "2" (with "0" indicating no signaling entity present), and the codewords distributed to define a ternary error-correcting code. Similarly, a third signaling entity may additionally be assigned as "3" to make a quaternary error-correcting code, etc.

As discussed above, in certain aspects, signaling entities are determined, e.g., to determine nucleic acid probes and/or to create codewords. In some cases, signaling entities within a sample may be determined, e.g., spatially, using a variety of techniques. In some embodiments, the signaling entities may be fluorescent, and techniques for determining fluorescence within a sample, such as fluorescence microscopy or confocal microscopy, may be used to spatially identify the positions of signaling entities within a cell. In some cases, the positions of entities within the sample may be determined in two or even three dimensions. In addition, in some embodiments, more than one signaling entity may be determined at a time (e.g., signaling entities with different colors or emissions), and/or sequentially.

In addition, in some embodiments, a confidence level for the identified nucleic acid target may be determined. For example, the confidence level may be determined using a ratio of the number of exact matches to the number of matches having one or more one-bit errors. In some cases, only matches having a confidence ratio greater than a certain value may be used. For instance, in certain embodiments, matches may be accepted only if the confidence ratio for the match is greater than about 0.01, greater than about 0.03, greater than about 0.05, greater than about 0.1, greater than about 0.3, greater than about 0.5, greater than about 1, greater than about 3, greater than about 5, greater than about 10, greater than about 30, greater than about 50, greater than about 100, greater than about 300, greater than about 500, greater than about 1000, or any other suitable value. In addition, in some embodiments, matches may be accepted only if the confidence ratio for the identified nucleic acid target is greater than an internal standard or false positive control by about 0.01, about 0.03, about

0.05, about 0.1, about 0.3, about 0.5, about 1, about 3, about 5, about 10, about 30, about 50, about 100, about 300, about 500, about 1000, or any other suitable value

In some embodiments, the spatial positions of the entities (and thus, nucleic acid probes that the entities may be associated with) may be determined at relatively high resolutions. For instance, the positions may be determined at spatial resolutions of better than about 100 micrometers, better than about 30 micrometers, better than about 10 micrometers, better than about 3 micrometers, better than about 1 micrometer, better than about 800 nm, better than about 600 nm, better than about 500 nm, better than about 400 nm, better than about 300 nm, better than about 200 nm, better than about 100 nm, better than about 90 nm, better than about 80 nm, better than about 70 nm, better than about 60 nm, better than about 50 nm, better than about 40 nm, better than about 30 nm, better than about 20 nm, or better than about 10 nm, etc.

There are a variety of techniques able to determine or image the spatial positions of entities optically, e.g., using fluorescence microscopy. In some cases, the spatial positions may be determined at super resolutions, or at resolutions better than the wavelength of light or the diffraction limit. Non-limiting examples include STORM (stochastic optical reconstruction microscopy), STED (stimulated emission depletion microscopy), NSOM (Near-field Scanning Optical Microscopy), 4Pi microscopy, SIM (Structured Illumination Microscopy), SMI (Spatially Modulated Illumination) microscopy, RESOLFT (Reversible Saturable Optically Linear Fluorescence Transition Microscopy), GSD (Ground State Depletion Microscopy), SSIM (Saturated Structured-Illumination Microscopy), SPDM (Spectral Precision Distance Microscopy), Photo-Activated Localization Microscopy (PALM), Fluorescence Photoactivation Localization Microscopy (FPALM), LIMON (3D Light Microscopical Nanosizing Microscopy), Super-resolution optical fluctuation imaging (SOFI), or the like. See, e.g., U.S. Pat. No. 7,838,302, issued November 23, 2010, entitled "Sub-Diffraction Limit Image Resolution and Other Imaging Techniques," by Zhuang, *et al.*; U.S. Pat. No. 8,564,792, issued October 22, 2013, entitled "Sub-diffraction Limit Image Resolution in Three Dimensions," by Zhuang, *et al.*; or Int. Pat. Apl. Pub. No. WO 2013/090360, published June 20, 2013, entitled "High Resolution Dual-Objective Microscopy," by Zhuang, *et al.*, each incorporated herein by reference in their entireties.

As an illustrative non-limiting example, in one set of embodiments, the sample may be imaged with a high numerical aperture, oil immersion objective with 100X magnification and light collected on an electron-multiplying CCD camera. In another example, the sample could be imaged with a high numerical aperture, oil immersion lens with 40X magnification

and light collected with a wide-field scientific CMOS camera. With different combinations of objectives and cameras, a single field of view may correspond to no less than 40 x 40 microns, 80 x 80 microns, 120 x 120 microns, 240 x 240 microns, 340 x 340 microns, or 500 x 500 microns, etc. in various non-limiting embodiments. Similarly, a single camera pixel may correspond, in some embodiments, to regions of the sample of no less than 80x80 nm, 120x120 nm, 160x160 nm, 240x240 nm, or 300x300 nm, etc. In another example, the sample may be imaged with a low numerical aperture, air lens with 10X magnification and light collected with a sCMOS camera. In additional embodiments, the sample may be optically sectioned by illuminating it via a single or multiple scanned diffraction limited foci generated either by scanning mirrors or a spinning disk and the collected passed through a single or multiple pinholes. In another embodiment, the sample may also be illuminated via thin sheet of light generated via any one of multiple methods known to those versed in the art.

In one embodiment, the sample may be illuminated by single Gaussian mode laser lines. In some embodiments, the illumination profiled may be flattened by passing these laser lines through a multimode fiber that is vibrated via piezo-electric or other mechanical means. In some embodiments, the illumination profile may be flattened by passing single-mode, Gaussian beams through a variety of refractive beam shapers, such as the piShaper or a series of stacked Powell lenses. In yet another set of embodiments, the Gaussian beams may be passed through a variety of different diffusing elements, such as ground glass or engineered diffusers, which may be spun in some cases at high speeds to remove residual laser speckle. In yet another embodiment, laser illumination may be passed through a series of lenslet arrays to produce overlapping images of the illumination that approximate a flat illumination field.

In some embodiments, the centroids of the spatial positions of the entities may be determined. For example, a centroid of a signaling entity may be determined within an image or series of images using image analysis algorithms known to those of ordinary skill in the art. In some cases, the algorithms may be selected to determine non-overlapping single emitters and/or partially overlapping single emitters in a sample. Non-limiting examples of suitable techniques include a maximum likelihood algorithm, a least squares algorithm, a Bayesian algorithm, a compressed sensing algorithm, or the like. Combinations of these techniques may also be used in some cases.

In addition, the signaling entity may be inactivated in some cases. For example, in some embodiments, a first secondary nucleic acid probe containing a signaling entity may be applied to a sample that can recognize a first read sequence, then the first secondary nucleic

acid probe can be inactivated before a second secondary nucleic acid probe is applied to the sample. If multiple signaling entities are used, the same or different techniques may be used to inactivate the signaling entities, and some or all of the multiple signaling entities may be inactivated, e.g., sequentially or simultaneously.

Inactivation may be caused by removal of the signaling entity (e.g., from the sample, or from the nucleic acid probe, etc.), and/or by chemically altering the signaling entity in some fashion, e.g., by photobleaching the signaling entity, bleaching or chemically altering the structure of the signaling entity, e.g., by reduction, etc.). For instance, in one set of embodiments, a fluorescent signaling entity may be inactivated by chemical or optical techniques such as oxidation, photobleaching, chemically bleaching, stringent washing or enzymatic digestion or reaction by exposure to an enzyme, dissociating the signaling entity from other components (e.g., a probe), chemical reaction of the signaling entity (e.g., to a reactant able to alter the structure of the signaling entity) or the like. For instance, bleaching may occur by exposure to oxygen, reducing agents, or the signaling entity could be chemically cleaved from the nucleic acid probe and washed away via fluid flow.

In some embodiments, various nucleic acid probes (including primary and/or secondary nucleic acid probes) may include one or more signaling entities. If more than one nucleic acid probe is used, the signaling entities may each by the same or different. In certain embodiments, a signaling entity is any entity able to emit light. For instance, in one embodiment, the signaling entity is fluorescent. In other embodiments, the signaling entity may be phosphorescent, radioactive, absorptive, etc. In some cases, the signaling entity is any entity that can be determined within a sample at relatively high resolutions, e.g., at resolutions better than the wavelength of visible light or the diffraction limit. The signaling entity may be, for example, a dye, a small molecule, a peptide or protein, or the like. The signaling entity may be a single molecule in some cases. If multiple secondary nucleic acid probes are used, the nucleic acid probes may comprise the same or different signaling entities.

Non-limiting examples of signaling entities include fluorescent entities (fluorophores) or phosphorescent entities, for example, cyanine dyes (e.g., Cy2, Cy3, Cy3B, Cy5, Cy5.5, Cy7, etc.), Alexa Fluor dyes, Atto dyes, photoswtichable dyes, photoactivatable dyes, fluorescent dyes, metal nanoparticles, semiconductor nanoparticles or "quantum dots", fluorescent proteins such as GFP (Green Fluorescent Protein), or photoactivabale fluorescent proteins, such as PAGFP, PSCFP, PSCFP2, Dendra, Dendra2, EosFP, tdEos, mEos2, mEos3, PAmCherry, PAtagRFP, mMaple, mMaple2, and mMaple3. Other suitable signaling entities

are known to those of ordinary skill in the art. See, e.g., U.S. Pat. No. 7,838,302 or U.S. Pat. Apl. Ser. No. 61/979,436, each incorporated herein by reference in its entirety.

In one set of embodiments, the signaling entity may be attached to an oligonucleotide sequence via a bond that can be cleaved to release the signaling entity. In one set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a cleavable bond, such as a photocleavable bond. Non-limiting examples of photocleavable bonds include, but are not limited to, 1-(2-nitrophenyl)ethyl, 2-nitrobenzyl, biotin phosphoramidite, acrylic phosphoramidite, diethylaminocoumarin, 1-(4,5-dimethoxy-2-nitrophenyl)ethyl, cyclo-dodecyl (dimethoxy-2-nitrophenyl)ethyl, 4-aminomethyl-3-nitrobenzyl, (4-nitro-3-(1-chlorocarbonyloxyethyl)phenyl)methyl-S-acetylthioic acid ester, (4-nitro-3-(1-thlorocarbonyloxyethyl)phenyl)methyl-3-(2-pyridyldithiopropionic acid) ester, 3-(4,4'-dimethoxytrityl)-1-(2-nitrophenyl)-propane-1,3-diol-[2- cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-trifluoroacetylcaproamidomethyl)phenyl]-ethyl-[2-cyano-ethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(4,4'-dimethoxytrityloxy)butyramidomethyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, 1-[2-nitro-5-(6-(N-(4,4'-dimethoxytrityl))-biotinamidocaproamido-methyl)phenyl]-ethyl-[2-cyanoethyl-(N,N-diisopropyl)]-phosphoramidite, or similar linkers. In another set of embodiments, the fluorophore may be conjugated to an oligonucleotide via a disulfide bond. The disulfide bond may be cleaved by a variety of reducing agents such as, but not limited to, dithiothreitol, dithioerythritol, beta-mercaptoethanol, sodium borohydride, thioredoxin, glutaredoxin, trypsinogen, hydrazine, diisobutylaluminum hydride, oxalic acid, formic acid, ascorbic acid, phosphorous acid, tin chloride, glutathione, thioglycolate, 2,3-dimercaptopropanol, 2-mercaptoethylamine, 2-aminoethanol, tris(2-carboxyethyl)phosphine, bis(2-mercaptoethyl) sulfone, N,N'-dimethyl-N,N'-bis(mercaptoacetyl)hydrazine, 3-mercaptoproptionate, dimethylformamide, thiopropyl-agarose, tri-*n*-butylphosphine, cysteine, iron sulfate, sodium sulfite, phosphite, hypophosphite, phosphorothioate, or the like, and/or combinations of any of these. In another embodiment, the fluorophore may be conjugated to an oligonucleotide via one or more phosphorothioate modified nucleotides in which the sulfur modification replaces the bridging and/or non-bridging oxygen. The fluorophore may be cleaved from the oligonucleotide, in certain embodiments, via addition of compounds such as but not limited to iodoethanol, iodine mixed in ethanol, silver nitrate, or mercury chloride. In yet another set of embodiments, the signaling entity may be chemically inactivated through reduction or oxidation. For example, in one embodiment, a chromophore such as Cy5 or Cy7 may be reduced using sodium borohydride to a stable, non-fluorescence state. In still another

set of embodiments, a fluorophore may be conjugated to an oligonucleotide via an azo bond, and the azo bond may be cleaved with 2-[(2-N-arylamino)phenylazo]pyridine. In yet another set of embodiments, a fluorophore may be conjugated to an oligonucleotide via a suitable nucleic acid segment that can be cleaved upon suitable exposure to DNAse, e.g., an exodeoxyribonuclease or an endodeoxyribonuclease. Examples include, but are not limited to, deoxyribonuclease I or deoxyribonuclease II. In one set of embodiments, the cleavage may occur via a restriction endonuclease. Non-limiting examples of potentially suitable restriction endonucleases include BamHI, BsrI, NotI, XmaI, PspAI, DpnI, MboI, MnlI, Eco57I, Ksp632I, DraIII, AhaII, SmaI, MluI, HpaI, ApaI, BclI, BstEII, TaqI, EcoRI, SacI, HindII, HaeII, DraII, Tsp509I, Sau3AI, PacI, etc. Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially. In yet another set of embodiments, a fluorophore may be conjugated to biotin, and the oligonucleotide conjugated to avidin or streptavidin. An interaction between biotin and avidin or streptavidin allows the fluorophore to be conjugated to the oligonucleotide, while sufficient exposure to an excess of addition, free biotin could "outcompete" the linkage and thereby cause cleavage to occur. In addition, in another set of embodiments, the probes may be removed using corresponding "toe-hold-probes," which comprise the same sequence as the probe, as well as an extra number of bases of homology to the encoding probes (e.g., 1-20 extra bases, for example, 5 extra bases). These probes may remove the labeled readout probe through a strand-displacement interaction.

As used herein, the term "light" generally refers to electromagnetic radiation, having any suitable wavelength (or equivalently, frequency). For instance, in some embodiments, the light may include wavelengths in the optical or visual range (for example, having a wavelength of between about 400 nm and about 700 nm, i.e., "visible light"), infrared wavelengths (for example, having a wavelength of between about 300 micrometers and 700 nm), ultraviolet wavelengths (for example, having a wavelength of between about 400 nm and about 10 nm), or the like. In certain cases, as discussed in detail below, more than one entity may be used, i.e., entities that are chemically different or distinct, for example, structurally. However, in other cases, the entities may be chemically identical or at least substantially chemically identical.

In one set of embodiments, the signaling entity is "switchable," i.e., the entity can be switched between two or more states, at least one of which emits light having a desired wavelength. In the other state(s), the entity may emit no light, or emit light at a different wavelength. For instance, an entity may be "activated" to a first state able to produce light

having a desired wavelength, and "deactivated" to a second state not able to emit light of the same wavelength. An entity is "photoactivatable" if it can be activated by incident light of a suitable wavelength. As a non-limiting example, Cy5, can be switched between a fluorescent and a dark state in a controlled and reversible manner by light of different wavelengths, i.e., 633 nm (or 642nm, 647nm, 656 nm) red light can switch or deactivate Cy5 to a stable dark state, while 405 nm green light can switch or activate the Cy5 back to the fluorescent state. In some cases, the entity can be reversibly switched between the two or more states, e.g., upon exposure to the proper stimuli. For example, a first stimuli (e.g., a first wavelength of light) may be used to activate the switchable entity, while a second stimuli (e.g., a second wavelength of light) may be used to deactivate the switchable entity, for instance, to a non-emitting state. Any suitable method may be used to activate the entity. For example, in one embodiment, incident light of a suitable wavelength may be used to activate the entity to emit light, i.e., the entity is "photoswitchable." Thus, the photoswitchable entity can be switched between different light-emitting or non-emitting states by incident light, e.g., of different wavelengths. The light may be monochromatic (e.g., produced using a laser) or polychromatic. In another embodiment, the entity may be activated upon stimulation by electric field and/or magnetic field. In other embodiments, the entity may be activated upon exposure to a suitable chemical environment, e.g., by adjusting the pH, or inducing a reversible chemical reaction involving the entity, etc. Similarly, any suitable method may be used to deactivate the entity, and the methods of activating and deactivating the entity need not be the same. For instance, the entity may be deactivated upon exposure to incident light of a suitable wavelength, or the entity may be deactivated by waiting a sufficient time.

Typically, a "switchable" entity can be identified by one of ordinary skill in the art by determining conditions under which an entity in a first state can emit light when exposed to an excitation wavelength, switching the entity from the first state to the second state, e.g., upon exposure to light of a switching wavelength, then showing that the entity, while in the second state can no longer emit light (or emits light at a much reduced intensity) when exposed to the excitation wavelength.

In one set of embodiments, as discussed, a switchable entity may be switched upon exposure to light. In some cases, the light used to activate the switchable entity may come from an external source, e.g., a light source such as a laser light source, another light-emitting entity proximate the switchable entity, etc. The second, light emitting entity, in some cases, may be a fluorescent entity, and in certain embodiments, the second, light-emitting entity may itself also be a switchable entity.

In some embodiments, the switchable entity includes a first, light-emitting portion (e.g., a fluorophore), and a second portion that activates or "switches" the first portion. For example, upon exposure to light, the second portion of the switchable entity may activate the first portion, causing the first portion to emit light. Examples of activator portions include, but are not limited to, Alexa Fluor 405 (Invitrogen), Alexa Fluor 488 (Invitrogen), Cy2 (GE Healthcare), Cy3 (GE Healthcare), Cy3B (GE Healthcare), Cy3.5 (GE Healthcare), or other suitable dyes. Examples of light-emitting portions include, but are not limited to, Cy5, Cy5.5 (GE Healthcare), Cy7 (GE Healthcare), Alexa Fluor 647 (Invitrogen), Alexa Fluor 680 (Invitrogen), Alexa Fluor 700 (Invitrogen), Alexa Fluor 750 (Invitrogen), Alexa Fluor 790 (Invitrogen), DiD, DiR, YOYO-3 (Invitrogen), YO-PRO-3 (Invitrogen), TOT-3 (Invitrogen), TO-PRO-3 (Invitrogen) or other suitable dyes. These may linked together, e.g., covalently, for example, directly, or through a linker, e.g., forming compounds such as, but not limited to, Cy5-Alexa Fluor 405, Cy5-Alexa Fluor 488, Cy5-Cy2, Cy5-Cy3, Cy5-Cy3.5, Cy5.5-Alexa Fluor 405, Cy5.5-Alexa Fluor 488, Cy5.5-Cy2, Cy5.5-Cy3, Cy5.5-Cy3.5, Cy7-Alexa Fluor 405, Cy7-Alexa Fluor 488, Cy7-Cy2, Cy7-Cy3, Cy7-Cy3.5, Alexa Fluor 647-Alexa Fluor 405, Alexa Fluor 647-Alexa Fluor 488, Alexa Fluor 647-Cy2, Alexa Fluor 647-Cy3, Alexa Fluor 647-Cy3.5, Alexa Fluor 750-Alexa Fluor 405, Alexa Fluor 750-Alexa Fluor 488, Alexa Fluor 750-Cy2, Alexa Fluor 750-Cy3, or Alexa Fluor 750-Cy3.5. Those of ordinary skill in the art will be aware of the structures of these and other compounds, many of which are available commercially. The portions may be linked via a covalent bond, or by a linker, such as those described in detail below. Other light-emitting or activator portions may include portions having two quaternized nitrogen atoms joined by a polymethine chain, where each nitrogen is independently part of a heteroaromatic moiety, such as pyrrole, imidazole, thiazole, pyridine, quinoine, indole, benzothiazole, etc., or part of a nonaromatic amine. In some cases, there may be 5, 6, 7, 8, 9, or more carbon atoms between the two nitrogen atoms.

In certain cases, the light-emitting portion and the activator portions, when isolated from each other, may each be fluorophores, i.e., entities that can emit light of a certain, emission wavelength when exposed to a stimulus, for example, an excitation wavelength. However, when a switchable entity is formed that comprises the first fluorophore and the second fluorophore, the first fluorophore forms a first, light-emitting portion and the second fluorophore forms an activator portion that switches that activates or "switches" the first portion in response to a stimulus. For example, the switchable entity may comprise a first fluorophore directly bonded to the second fluorophore, or the first and second entity may be

connected via a linker or a common entity. Whether a pair of light-emitting portion and activator portion produces a suitable switchable entity can be tested by methods known to those of ordinary skills in the art. For example, light of various wavelength can be used to stimulate the pair and emission light from the light-emitting portion can be measured to determined wither the pair makes a suitable switch.

As a non-limiting example, Cy3 and Cy5 may be linked together to form such an entity. In this example, Cy3 is an activator portion that is able to activate Cy5, the light-emission portion. Thus, light at or near the absorption maximum (e.g., near 532 nm light for Cy3) of the activation or second portion of the entity may cause that portion to activate the first, light-emitting portion, thereby causing the first portion to emit light (e.g., near 647 nm for Cy5). See, e.g., U.S. Pat. No. 7,838,302, incorporated herein by reference in its entirety. In some cases, the first, light-emitting portion can subsequently be deactivated by any suitable technique (e.g., by directing 647 nm red light to the Cy5 portion of the molecule).

Other non-limiting examples of potentially suitable activator portions include 1,5 IAEDANS, 1,8-ANS, 4-Methylumbelliferone, 5-carboxy-2,7-dichlorofluorescein, 5-Carboxyfluorescein (5-FAM), 5-Carboxynapthofluorescein, 5-Carboxytetramethylrhodamine (5-TAMRA), 5-FAM (5-Carboxyfluorescein), 5-HAT (Hydroxy Tryptamine), 5-Hydroxy Tryptamine (HAT), 5-ROX (carboxy-X-rhodamine), 5-TAMRA (5-Carboxytetramethylrhodamine), 6-Carboxyrhodamine 6G, 6-CR 6G, 6-JOE, 7-Amino-4-methylcoumarin, 7-Aminoactinomycin D (7-AAD), 7-Hydroxy-4-methylcoumarin, 9-Amino-6-chloro-2-methoxyacridine, ABQ, Acid Fuchsin, ACMA (9-Amino-6-chloro-2-methoxyacridine), Acridine Orange, Acridine Red, Acridine Yellow, Acriflavin, Acriflavin Feulgen SITSA, Alexa Fluor 350, Alexa Fluor 405, Alexa Fluor 430, Alexa Fluor 488, Alexa Fluor 500, Alexa Fluor 514, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 610, Alexa Fluor 633, Alexa Fluor 635, Alizarin Complexon, Alizarin Red, AMC, AMCA-S, AMCA (Aminomethylcoumarin), AMCA-X, Aminoactinomycin D, Aminocoumarin, Aminomethylcoumarin (AMCA), Anilin Blue, Anthrocyl stearate, APTRA-BTC, APTS, Astrazon Brilliant Red 4G, Astrazon Orange R, Astrazon Red 6B, Astrazon Yellow 7 GLL, Atabrine, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 520, ATTO 532, ATTO 550, ATTO 565, ATTO 590, ATTO 594, ATTO 610, ATTO 611X, ATTO 620, ATTO 633, ATTO 635, ATTO 647, ATTO 647N, ATTO 655, ATTO 680, ATTO 700, ATTO 725, ATTO 740, ATTO-TAG CBQCA, ATTO-TAG FQ, Auramine, Aurophosphine G, Aurophosphine, BAO 9 (Bisaminophenyloxadiazole), BCECF (high pH), BCECF (low pH), Berberine Sulphate,

Bimane, Bisbenzamide, Bisbenzimide (Hoechst), bis-BTC, Blancophor FFG, Blancophor SV, BOBO -1, BOBO -3, Bodipy 492/515, Bodipy 493/503, Bodipy 500/510, Bodipy 505/515, Bodipy 530/550, Bodipy 542/563, Bodipy 558/568, Bodipy 564/570, Bodipy 576/589, Bodipy 581/591, Bodipy 630/650-X, Bodipy 650/665-X, Bodipy 665/676, Bodipy Fl, Bodipy FL ATP, Bodipy Fl-Ceramide, Bodipy R6G, Bodipy TMR, Bodipy TMR-X conjugate, Bodipy TMR-X, SE, Bodipy TR, Bodipy TR ATP, Bodipy TR-X SE, BO-PRO -1, BO-PRO -3, Brilliant Sulphoflavin FF, BTC, BTC-5N, Calcein, Calcein Blue, Calcium Crimson, Calcium Green, Calcium Green-1 Ca^{2+} Dye, Calcium Green-2 Ca^{2+}, Calcium Green-5N Ca^{2+}, Calcium Green-C18 Ca^{2+}, Calcium Orange, Calcofluor White, Carboxy-X-rhodamine (5-ROX), Cascade Blue, Cascade Yellow, Catecholamine, CCF2 (GeneBlazer), CFDA, Chromomycin A, Chromomycin A, CL-NERF, CMFDA, Coumarin Phalloidin, CPM Methylcoumarin, CTC, CTC Formazan, Cy2, Cy3.1 8, Cy3.5, Cy3, Cy5.1 8, cyclic AMP Fluorosensor (FiCRhR), Dabcyl, Dansyl, Dansyl Amine, Dansyl Cadaverine, Dansyl Chloride, Dansyl DHPE, Dansyl fluoride, DAPI, Dapoxyl, Dapoxyl 2, Dapoxyl 3' DCFDA, DCFH (Dichlorodihydrofluorescein Diacetate), DDAO, DHR (Dihydorhodamine 123), Di-4-ANEPPS, Di-8-ANEPPS (non-ratio), DiA (4-Di-16-ASP), Dichlorodihydrofluorescein Diacetate (DCFH), DiD - Lipophilic Tracer, DiD (DiIC18(5)), DIDS, Dihydorhodamine 123 (DHR), DiI (DiIC18(3)), Dinitrophenol, DiO (DiOC18(3)), DiR, DiR (DiIC18(7)), DM-NERF (high pH), DNP, Dopamine, DTAF, DY-630-NHS, DY-635-NHS, DyLight 405, DyLight 488, DyLight 549, DyLight 633, DyLight 649, DyLight 680, DyLight 800, ELF 97, Eosin, Erythrosin, Erythrosin ITC, Ethidium Bromide, Ethidium homodimer -1 (EthD-1), Euchrysin, EukoLight, Europium (III) chloride, Fast Blue, FDA, Feulgen (Pararosaniline), FIF (Formaldehyd Induced Fluorescence), FITC, Flazo Orange, Fluo-3, Fluo-4, Fluorescein (FITC), Fluorescein Diacetate, Fluoro-Emerald, Fluoro-Gold (Hydroxystilbamidine), Fluor-Ruby, FluorX, FM 1-43, FM 4-46, Fura Red (high pH), Fura Red/Fluo-3, Fura-2, Fura-2/BCECF, Genacryl Brilliant Red B, Genacryl Brilliant Yellow 10GF, Genacryl Pink 3G, Genacryl Yellow 5GF, GeneBlazer (CCF2), Gloxalic Acid, Granular blue, Haematoporphyrin, Hoechst 33258, Hoechst 33342, Hoechst 34580, HPTS, Hydroxycoumarin, Hydroxystilbamidine (FluoroGold), Hydroxytryptamine, Indo-1, high calcium, Indo-1, low calcium, Indodicarbocyanine (DiD), Indotricarbocyanine (DiR), Intrawhite Cf, JC-1, JO-JO-1, JO-PRO-1, LaserPro, Laurodan, LDS 751 (DNA), LDS 751 (RNA), Leucophor PAF, Leucophor SF, Leucophor WS, Lissamine Rhodamine, Lissamine Rhodamine B, Calcein/Ethidium homodimer, LOLO-1, LO-PRO-1, Lucifer Yellow, Lyso Tracker Blue, Lyso Tracker Blue-White, Lyso Tracker Green, Lyso Tracker Red, Lyso

Tracker Yellow, LysoSensor Blue, LysoSensor Green, LysoSensor Yellow/Blue, Mag Green, Magdala Red (Phloxin B), Mag-Fura Red, Mag-Fura-2, Mag-Fura-5, Mag-Indo-1, Magnesium Green, Magnesium Orange, Malachite Green, Marina Blue, Maxilon Brilliant Flavin 10 GFF, Maxilon Brilliant Flavin 8 GFF, Merocyanin, Methoxycoumarin, Mitotracker Green FM, Mitotracker Orange, Mitotracker Red, Mitramycin, Monobromobimane, Monobromobimane (mBBr-GSH), Monochlorobimane, MPS (Methyl Green Pyronine Stilbene), NBD, NBD Amine, Nile Red, Nitrobenzoxadidole, Noradrenaline, Nuclear Fast Red, Nuclear Yellow, Nylosan Brilliant Iavin E8G, Oregon Green, Oregon Green 488-X, Oregon Green, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, Pararosaniline (Feulgen), PBFI, Phloxin B (Magdala Red), Phorwite AR, Phorwite BKL, Phorwite Rev, Phorwite RPA, Phosphine 3R, PKH26 (Sigma), PKH67, PMIA, Pontochrome Blue Black, POPO-1, POPO-3, PO-PRO-1, PO-PRO-3, Primuline, Procion Yellow, Propidium Iodid (PI), PyMPO, Pyrene, Pyronine, Pyronine B, Pyrozal Brilliant Flavin 7GF, QSY 7, Quinacrine Mustard, Resorufin, RH 414, Rhod-2, Rhodamine, Rhodamine 110, Rhodamine 123, Rhodamine 5 GLD, Rhodamine 6G, Rhodamine B, Rhodamine B 200, Rhodamine B extra, Rhodamine BB, Rhodamine BG, Rhodamine Green, Rhodamine Phallicidine, Rhodamine Phalloidine, Rhodamine Red, Rhodamine WT, Rose Bengal, S65A, S65C, S65L, S65T, SBFI, Serotonin, Sevron Brilliant Red 2B, Sevron Brilliant Red 4G, Sevron Brilliant Red B, Sevron Orange, Sevron Yellow L, SITS, SITS (Primuline), SITS (Stilbene Isothiosulphonic Acid), SNAFL calcein, SNAFL-1, SNAFL-2, SNARF calcein, SNARF1, Sodium Green, SpectrumAqua, SpectrumGreen, SpectrumOrange, Spectrum Red, SPQ (6-methoxy-N-(3-sulfopropyl)quinolinium), Stilbene, Sulphorhodamine B can C, Sulphorhodamine Extra, SYTO 11, SYTO 12, SYTO 13, SYTO 14, SYTO 15, SYTO 16, SYTO 17, SYTO 18, SYTO 20, SYTO 21, SYTO 22, SYTO 23, SYTO 24, SYTO 25, SYTO 40, SYTO 41, SYTO 42, SYTO 43, SYTO 44, SYTO 45, SYTO 59, SYTO 60, SYTO 61, SYTO 62, SYTO 63, SYTO 64, SYTO 80, SYTO 81, SYTO 82, SYTO 83, SYTO 84, SYTO 85, SYTOX Blue, SYTOX Green, SYTOX Orange, Tetracycline, Tetramethylrhodamine (TAMRA), Texas Red, Texas Red-X conjugate, Thiadicarbocyanine (DiSC3), Thiazine Red R, Thiazole Orange, Thioflavin 5, Thioflavin S, Thioflavin TCN, Thiolyte, Thiozole Orange, Tinopol CBS (Calcofluor White), TMR, TO-PRO-1, TO-PRO-3, TO-PRO-5, TOTO-1, TOTO-3, TRITC (tetramethylrodamine isothiocyanate), True Blue, TruRed, Ultralite, Uranine B, Uvitex SFC, WW 781, X-Rhodamine, XRITC, Xylene Orange, Y66F, Y66H, Y66W, YO-PRO-1, YO-PRO-3, YOYO-1, YOYO-3, SYBR Green, Thiazole orange (interchelating dyes), or combinations thereof.