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.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/379,544	10/11/2022	11466308	080015-028420US-1249377	2092

20350 7590 09/21/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 0 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Patents Stakeholder Experience (OPSE), Stakeholder Support Division (SSD) at (571)-272-4200.

INVENTOR(s) (Please see PAIR WEB site http://pair.uspto.gov for additional inventors):

Yuk-Ming Dennis Lo, Homantin, CHINA;
Rossa Wai Kwun Chiu, Shatin, CHINA;
Kwan Chee Chan, Mei Foo Sun Cheun, CHINA;
Peiyong Jiang, Pak Shek Kok, CHINA;
Suk Hang Cheng, Fanling, CHINA;

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

The Chinese University of Hong Kong, Shatin, HONG KONG;

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 USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.

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.
17/379,544	07/19/2021	Yuk-Ming Dennis Lo	080015-028420US-1249377	2092

20350 7590 08/17/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

EXAMINER
GODDARD, LAURA B

ART UNIT	PAPER NUMBER
1642	

NOTIFICATION DATE	DELIVERY MODE
08/17/2022	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):

KTSDocketing2@kilpatrick.foundationip.com
ipefiling@kilpatricktownsend.com

Notice of Allowability	Application No. 17/379,544	Applicant(s) Lo et al.	
	Examiner LAURA B GODDARD	Art Unit 1642	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 8/9/2022.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ 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.

3. ☑ The allowed claim(s) is/are 1-13 and 18-27 . As a result of the allowed claim(s), 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.**

4. ☐ 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)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/Laura B Goddard/
Primary Examiner, Art Unit 1642

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-028420US-1249377

U.S. PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code [1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear

U.S. PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code [1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T[5]

NON-PATENT LITERATURE DOCUMENTS

Examiner 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.	T[5]
	1.	Communication pursuant to Article 94(3) EPC dated August 4, 2022 in EP Patent Application No. 20853612.8. 6 pages.	

EXAMINER SIGNATURE

Examiner Signature	/Laura B Goddard/	Date Considered	08/13/2022	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document.
[4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

Doc Code: MFEE.C.AD
Document Description: Maintenance Fee Address Change

PTO/AIA/47 (11-21)
Approved for use through 11/30/2024. OMB 0651-0016
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number.

"FEE ADDRESS" INDICATION FORM

Address to:
Mail Stop Maintenance Fee
Director of the United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

-OR-

Fax to:
571-273-6500

INSTRUCTIONS: The issue fee must have been paid for the application(s) listed on this form. In addition, only an address represented by a Customer Number can be established as the fee address for maintenance fee purposes (hereafter, fee address). A fee address should be established when correspondence related to maintenance fees should be mailed to a different address than the correspondence address for the application.

When to check the first box below: If you have a Customer Number to represent the fee address.

When to check the second box below: If you have no Customer Number representing the desired fee address, in which case a completed Request for Customer Number (PTO/SB/125) must be attached to this form. For more information on Customer Numbers, see the Manual of Patent Examining Procedure (MPEP) § 403.

For the following listed application(s), please recognize as the "Fee Address" under the provisions of 37 CFR 1.363 the address associated with:

☑ Customer Number: 00197

OR

☐ The attached Request for Customer Number form (PTO/SB/125)

PATENT NUMBER (if known)	APPLICATION NUMBER
	17/379,544

Completed by (check one):

☐ Applicant/Inventor

☑ Attorney or Agent of record 71,659 (Reg. No.)

☐ Assignee of record of the entire interest. See 37 CFR 3.71. Statement under 37 CFR 3.73(c) (Form PTO/AIA/96) is enclosed or was filed on _____ (Date)

/David D. Hsu/
Signature

David D. Hsu
Typed or printed name

303-571-4000
Requester's telephone number

8/9/2022
Date

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. If the assignee is a juristic entity, this form must be signed by a patent practitioner (attorney or agent) of record. Submit multiple forms if more than one signature is required, see below*.

☑ *Total of 1 forms are submitted

A Federal agency may not conduct or sponsor, and a person is not required to respond to, nor shall a person be subject to a penalty for failure to comply with an information collection subject to the requirements of the Paperwork Reduction Act of 1995, unless the information collection has a currently valid OMB Control Number. The OMB Control Number for this information collection is 0651-0016. Public burden for this form is estimated to average 5 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the information collection. Send comments regarding this burden estimate or any other aspect of this information collection, including suggestions for reducing this burden to the Chief Administrative Officer, United States Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450 or email InformationCollection@uspto.gov. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. The United States Patent and Trademark Office (USPTO) collects this information under authority of 5 CFR 1.290. The information in this system of records is used to manage all records of applicant including name, citizenship, residence, post office address and other information with respect to inventors and their legal representatives pertaining to the applicant's activities in connection with the invention for which a patent is sought. This information is protected from disclosure to third parties in accordance with the Privacy Act.

However, routine uses of this information may include disclosure to the following: to law enforcement and investigation in the event that the system of records indicates a violation or potential violation of law; to a Federal, state, local, or international agency, in response to its request; to an agency, organization, or individual for the purpose of performing audit or oversight operations as authorized by law; to non-federal personnel under contract to the agency; to a court for adjudication and litigation; to the Department of Justice for Freedom of Information Act (FOIA) assistance; to members of congress working on behalf of an individual; to the Office of Personnel Management (OPM) for personnel research purposes; to National Archives and Records Administration for inspection of records; and to the Office of Management and Budget (OMB)for legislative coordination and clearance. Failure to provide any part of the requested information may result in an inability to process requests for access and information. The applicable Privacy Act System of Records Notice for this information is COMMERCE/PAT-TM-7 Patent Application Files, available at Federal Register / Vol. 78, No. 61 / Friday, March 29, 2013 /Notices 19243. https://www.govinfo.gov/content/pkg/FR-2013-03-29pdf/2013-07341.pdf

Electronic Patent Application Fee Transmittal

Application Number:	17379544
Filing Date:	19-Jul-2021
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Filer:	David Dah-Wei Hsu/Heather Orosco
Attorney Docket Number:	080015-028420US-1249377

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	1501	1	1200	1200

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**1200**

Electronic Acknowledgement Receipt

EFS ID:	46373934
Application Number:	17379544
International Application Number:	
Confirmation Number:	2092
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Customer Number:	20350
Filer:	David Dah-Wei Hsu/Heather Orosco
Filer Authorized By:	David Dah-Wei Hsu
Attorney Docket Number:	080015-028420US-1249377
Receipt Date:	09-AUG-2022
Filing Date:	19-JUL-2021
Time Stamp:	15:35:22
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1200
RAM confirmation Number	E202289F36073940
Deposit Account	201430
Authorized User	Heather Orosco

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.16 (National application filing, search, and examination fees)

37 CFR 1.17 (Patent application and reexamination processing fees)

37 CFR 1.19 (Document supply fees)

37 CFR 1.20 (Post Issuance fees)

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	Issue Fee Payment (PTO-85B)	1249377_Issue_Fee_Transmittal.pdf	1644338 f2c56b4b5a3042272c2847914717f7eda22c003f	no	1
Warnings:					
Information:					
2	Maintenance Fee Address Change	1249377_Fee_Address_Indication_Form.pdf	273852 73200778d460406292c9d0efbbc284ac02373e3d	no	2
Warnings:					
Information:					
3	Fee Worksheet (SB06)	fee-info.pdf	37935 b2291eb394d254fbcd686a5eebfa22930d14dceb	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	1956125		

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.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

20350 7590 05/10/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

Heather A. Orosco	(Typed or printed name)
/Heather A. Orosco/ via EFS-Web	(Signature)
August 9, 2022	(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/379,544	07/19/2021	Yuk-Ming Dennis Lo	080015-028420US-1249377	2092

TITLE OF INVENTION: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1200	$0.00	$0.00	$1200	08/10/2022

EXAMINER	ART UNIT	CLASS-SUBCLASS
GODDARD, LAURA B	1642	435-006110

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/AIA/122 or PTO/SB/122) attached.

☑ "Fee Address" indication (or "Fee Address" Indication form PTO/AIA/47 or PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 Kilpatrick Townsend & Stockton LLP
2
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE	(B) RESIDENCE: (CITY and STATE OR COUNTRY)
The Chinese University of Hong Kong	Shatin, New Territories, Hong Kong

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☑ Corporation or other private group entity ☐ Government

4a. Fees submitted: ☑ Issue Fee ☐ Publication Fee (if required) ☐ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

☑ Electronic Payment via EFS-Web ☐ Enclosed check ☐ Non-electronic payment by credit card (Attach form PTO-2038)

☑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. 20-1430

5. **Change in Entity Status** (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29

☐ Applicant asserting small entity status. See 37 CFR 1.27

☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /David D. Hsu/ Date 8/9/2022

Typed or printed name David D. Hsu Registration No. 71,659

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-028420US-1249377

U.S. PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code [1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear

U.S. PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code [1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T[5]

NON-PATENT LITERATURE DOCUMENTS

Examiner 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.	T[5]
	1.	Communication pursuant to Article 94(3) EPC dated August 4, 2022 in EP Patent Application No. 20853612.8. 6 pages.	

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document.
[4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

Electronic Patent Application Fee Transmittal

Application Number:	17379544
Filing Date:	19-Jul-2021
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Filer:	David Dah-Wei Hsu/Heather Orosco
Attorney Docket Number:	080015-028420US-1249377

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	1806	1	260	260
		Total in USD ($)		**260**

Electronic Acknowledgement Receipt

EFS ID:	46372078
Application Number:	17379544
International Application Number:	
Confirmation Number:	2092
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Customer Number:	20350
Filer:	David Dah-Wei Hsu/Heather Orosco
Filer Authorized By:	David Dah-Wei Hsu
Attorney Docket Number:	080015-028420US-1249377
Receipt Date:	09-AUG-2022
Filing Date:	19-JUL-2021
Time Stamp:	13:59:24
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$260
RAM confirmation Number	E202289E00051976
Deposit Account	201430
Authorized User	Heather Orosco

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

37 CFR 1.16 (National application filing, search, and examination fees)

37 CFR 1.17 (Patent application and reexamination processing fees)

37 CFR 1.19 (Document supply fees)

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		1249377_Supplemental_IDS.pdf	207922 c9c40a5a913d83636de1a7cc75941230038b2113	yes	3
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Transmittal Letter		1	2	
	Information Disclosure Statement (IDS) Form (SB08)		3	3	
Warnings:					
Information:					
2	Non Patent Literature	REF_Comm_pursuant_to_Article_94_3_EPC_in_EP20853612_8_dated_2022-08-04.pdf	229303 c87172de0af8ad8c47390834e6327254f5ab628e	no	6
Warnings:					
Information:					
3	Fee Worksheet (SB06)	fee-info.pdf	38084 c4fb49df9c9b5bfe9da3cd67b8b9790cc0ce115b	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	475309		

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.

I hereby certify that this correspondence is being filed via EFS-Web with the United States Patent and Trademark Office on August 9, 2022.

KILPATRICK TOWNSEND & STOCKTON LLP

By: /Heather A. Orosco/
Heather A. Orosco

PATENT
Attorney Docket No.: 080015-028420US-1249377
Client Ref. No.: 19/MED/901 US Con1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Confirmation No.: 2092
Yuk-Ming Dennis Lo	Examiner: Laura B. Goddard
Application No.: 17/379,544	Technology Center/Art Unit: 1642
Filed: July 19, 2021	
For: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS	SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT
Customer No.: 20350	

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

Commissioner:

The references cited on attached form PTO/SB/08A are being called to the attention of the Examiner. In accordance with the provisions of 37 CFR §1.98(a)(2), copies of any cited U.S. Patent and U.S. Patent Application Publications are not provided. Copies of the remaining cited references are provided.

It is respectfully requested that the cited references be expressly considered during the prosecution of this application, and the references be made of record therein and appear among the "references cited" on any patent to issue therefrom.

As provided for by 37 CFR §1.97(g) and (h), no inference should be made that the information and references cited are prior art merely because they are in this statement and no

representation is being made that a search has been conducted or that this statement encompasses all the possible relevant information.

This IDS is being filed after an action closing prosecution, but before the payment of the Issue Fee.

I hereby certify under 37 CFR §1.97(e)(1) that each item of information contained in this Information Disclosure Statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this statement.

The Commissioner is authorized to charge the fee set forth in §1.17(p) to the firm's credit card for consideration of this paper. The Commissioner is authorized to charge any additional fee due to Deposit Account No. 20-1430.

Respectfully submitted,

/David D. Hsu/

David D. Hsu
Registration No. 71659

KILPATRICK TOWNSEND & STOCKTON LLP

Attachment

76363530V.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

NOTICE OF ALLOWANCE AND FEE(S) DUE

20350 7590 05/10/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

EXAMINER
GODDARD, LAURA B

ART UNIT	PAPER NUMBER
1642	

DATE MAILED: 05/10/2022

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/379,544	07/19/2021	Yuk-Ming Dennis Lo	080015-028420US-1249377	2092

TITLE OF INVENTION: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1200	$0.00	$0.00	$1200	08/10/2022

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Maintenance fees are due in utility patents issuing on applications filed on or after Dec. 12, 1980. It is patentee's responsibility to ensure timely payment of maintenance fees when due. More information is available at www.uspto.gov/PatentMaintenanceFees.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), by mail or fax, or via EFS-Web.

By mail, send to: Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

By fax, send to: (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

20350 7590 05/10/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being transmitted to the USPTO via EFS-Web or by facsimile to (571) 273-2885, on the date below.

(Typed or printed name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
17/379,544	07/19/2021	Yuk-Ming Dennis Lo	080015-028420US-1249377	2092

TITLE OF INVENTION: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$1200	$0.00	$0.00	$1200	08/10/2022

EXAMINER	ART UNIT	CLASS-SUBCLASS
GODDARD, LAURA B	1642	435-006110

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/AIA/122 or PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/AIA/47 or PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document must have been previously recorded, or filed for recordation, as set forth in 37 CFR 3.11 and 37 CFR 3.81(a). Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. Fees submitted: ❑ Issue Fee ❑ Publication Fee (if required) ❑ Advance Order - # of Copies _____

4b. Method of Payment: *(Please first reapply any previously paid fee shown above)*

❑ Electronic Payment via EFS-Web ❑ Enclosed check ❑ Non-electronic payment by credit card (Attach form PTO-2038)

❑ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment to Deposit Account No. _____

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

❑ Applicant asserting small entity status. See 37 CFR 1.27

❑ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

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.
17/379,544	07/19/2021	Yuk-Ming Dennis Lo	080015-028420US-1249377	2092

20350 7590 05/10/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

EXAMINER	
GODDARD, LAURA B	

ART UNIT	PAPER NUMBER
1642	

DATE MAILED: 05/10/2022

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. 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.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 17/379,544	Applicant(s) Lo et al.	
	Examiner LAURA B GODDARD	Art Unit 1642	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 4/27/2022.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ 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.

3. ☑ The allowed claim(s) is/are 1-13 and 18-27 . As a result of the allowed claim(s), 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.**

4. ☐ 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)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/Laura B Goddard/
Primary Examiner, Art Unit 1642

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/379,544	Lo et al.
	Examiner	**Art Unit**
	LAURA B GODDARD	1642

CPC						
Symbol					**Type**	**Version**
C12Q	/	1	/	6816	F	2013-01-01
C12N	/	9	/	22	I	2013-01-01
G16B	/	20	/	00	I	2019-02-01
G16B	/	30	/	00	I	2019-02-01
C12Q	/	1	/	6851	I	2013-01-01

CPC Combination Sets							
Symbol					**Type**	**Set**	**Ranking**	**Version**
	/		/					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	23	
/Laura B Goddard/ Primary Examiner, Art Unit 1642	03 May 2022	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	none

U.S. Patent and Trademark Office — Part of Paper No.: 20220503

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/379,544	Lo et al.
	Examiner	**Art Unit**
	LAURA B GODDARD	1642

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q	/	1	/	6816
G16B	/	20	/	00
G16B	/	30	/	00
C12Q	/	1	/	6851

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

NONE (Assistant Examiner) (Date)	**Total Claims Allowed:** 23	
/Laura B Goddard/ Primary Examiner, Art Unit 1642 (Primary Examiner) 03 May 2022 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure none

U.S. Patent and Trademark Office

Part of Paper No.: 20220503

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/379,544	Lo et al.
	Examiner	**Art Unit**
	LAURA B GODDARD	1642

☑ Claims renumbered in the same order as presented by applicant ☐ CPA ☑ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original

NONE (Assistant Examiner) (Date)	**Total Claims Allowed:** 23	
/Laura B Goddard/ Primary Examiner, Art Unit 1642 (Primary Examiner)	03 May 2022 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure none

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/379,544	Lo et al.
	Examiner	Art Unit
	LAURA B GODDARD	1642

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner
		LBG

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
WEST	01/31/2022	LBG
STN	01/31/2022	LBG
Review prosecution 16/995,607	01/31/2022	LBG
WEST	05/03/2022	LBG

/Laura B Goddard/ Primary Examiner, Art Unit 1642	

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/379,544	Lo et al.
	Examiner	Art Unit
	LAURA B GODDARD	1642

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**
C12Q	1/6816	05/03/2022	LBG
C12Q	1/6851	05/03/2022	LBG
G16B	20/00	05/03/2022	LBG
G16B	30/00	05/03/2022	LBG

/Laura B Goddard/ Primary Examiner, Art Unit 1642	

Bibliographic Data

Application No: 17/379,544

Foreign Priority claimed: ○ Yes ◉ No

35 USC 119 (a-d) conditions met: ☐ Yes ☑ No ☐ Met After Allowance

Verified and Acknowledged: /Laura B Goddard/ Examiner's Signature | Initials

Title: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
07/19/2021	435	1642	080015-028420US-1249377
RULE			

APPLICANTS

The Chinese University of Hong Kong, Shatin, HONG KONG

INVENTORS

Yuk-Ming Dennis Lo, Shatin, New Territories, HONG KONG

Rossa Wai Kwun Chiu, Shatin, New Territories, HONG KONG

Kwan Chee Chan, Mei Foo Sun Cheun, CHINA

Peiyong Jiang, Pak Shek Kok, CHINA

Suk Hang Cheng, Fanling, CHINA

Wenlei Peng, Shatin, CHINA

On Yee Tse, Fanling, CHINA

CONTINUING DATA

This application is a CON of 16995607 08/17/2020 PAT 11091794

16995607 has PRO of 63051210 07/13/2020

16995607 has PRO of 63019790 05/04/2020

16995607 has PRO of 62991891 03/19/2020

16995607 has PRO of 62970586 02/05/2020

16995607 has PRO of 62887987 08/16/2019

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

08/13/2021

STATE OR COUNTRY

HONG KONG

ADDRESS

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309
UNITED STATES

FILING FEE RECEIVED

$9,460

WEST Search History

Hide Items | Restore | Clear | Close

DATE: Tuesday, May 03, 2022

Hide?	Set Name	Query	Hit Count
DB=PGPB,USPT; PLUR=YES; OP=OR			
☐	L3	L2 AND (interpulse pulse duration)[CLM]	3
☐	L2	L1 AND (fetal fetus maternal pregnant mother)[CLM]	345
☐	L1	("single molecule real time sequencing" SMRT) AND (fetal fetus maternal mother)	2567

END OF SEARCH HISTORY

WEST Search History

Hide Items | Restore | Clear | Close

DATE: Monday, January 31, 2022

Hide?	Set Name	Query	Hit Count
DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI; PLUR=YES; OP=OR			
☐	L17	L16 AND ((pulse interpulse peak) NEAR (width distance))	5
☐	L16	L2 AND "Chinese University of Hong Kong"	26
☐	L15	L14 AND (fetal fetus maternal pregnant mother)	497
☐	L14	(C12Q1/6816 or C12N9/22 or G16B20/00 or G16B30/00 or C12Q1/6851).cpc. AND (SMRT "single molecule real time sequencing")	1018
☐	L13	L12 AND width[clm]	1
☐	L12	L8 AND (width position)[CLM]	34
☐	L11	L8 AND (interpulse pulse duration)[CLM]	0
☐	L10	L8 AND (position AND (distance width) AND (interpulse pulse duration))[CLM]	0
☐	L9	L4 AND (fetal fetus maternal pregnant mother)[CLM]	20
☐	L8	L3 AND (fetal fetus maternal pregnant mother)[CLM]	296
☐	L7	L5 AND Cas9	31
☐	L6	L5 AND interpulse	13
☐	L5	L4 AND (methylation methylated)	73
☐	L4	L2 AND position AND width AND pulse	115
☐	L3	L2 AND @ay<=2019	2248
☐	L2	("single molecule real time sequencing" SMRT) AND (fetal fetus maternal pregnant mother)	2543
☐	L1	("single molecule real time sequencing" SMRT) AND (fetal fetus maternal mother)	2493

END OF SEARCH HISTORY

Doc Code: DIST.E.FILE Document Description: Electronic Terminal Disclaimer - Filed	PTO/SB/26 U.S. Patent and Trademark Office Department of Commerce

Electronic Petition Request	**TERMINAL DISCLAIMER TO OBVIATE A DOUBLE PATENTING REJECTION OVER A "PRIOR" PATENT**
Application Number	17379544
Filing Date	19-Jul-2021
First Named Inventor	Yuk-Ming Lo
Attorney Docket Number	080015-028420US-1249377
Title of Invention	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS

☒ Filing of terminal disclaimer does not obviate requirement for response under 37 CFR 1.111 to outstanding Office Action

☒ This electronic Terminal Disclaimer is not being used for a Joint Research Agreement.

Owner	Percent Interest
The Chinese University of Hong Kong	100%

The owner(s) with percent interest listed above in the instant application hereby disclaims, except as provided below, the terminal part of the statutory term of any patent granted on the instant application which would extend beyond the expiration date of the full statutory term of prior patent number(s)

11091794

as the term of said prior patent is presently shortened by any terminal disclaimer. The owner hereby agrees that any patent so granted on the instant application shall be enforceable only for and during such period that it and the prior patent are commonly owned. This agreement runs with any patent granted on the instant application and is binding upon the grantee, its successors or assigns.

In making the above disclaimer, the owner does not disclaim the terminal part of the term of any patent granted on the instant application that would extend to the expiration date of the full statutory term of the prior patent, "as the term of said prior patent is presently shortened by any terminal disclaimer," in the event that said prior patent later:
- expires for failure to pay a maintenance fee;
- is held unenforceable;
- is found invalid by a court of competent jurisdiction;
- is statutorily disclaimed in whole or terminally disclaimed under 37 CFR 1.321;
- has all claims canceled by a reexamination certificate;
- is reissued; or
- is in any manner terminated prior to the expiration of its full statutory term as presently shortened by any terminal disclaimer.

◉ Terminal disclaimer fee under 37 CFR 1.20(d) is included with Electronic Terminal Disclaimer request.

○ I certify, in accordance with 37 CFR 1.4(d)(4), that the terminal disclaimer fee under 37 CFR 1.20(d) required for this terminal disclaimer has already been paid in the above-identified application.
Applicant claims the following fee status: ○ Small Entity ○ Micro Entity ◉ Regular Undiscounted I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.
THIS PORTION MUST BE COMPLETED BY THE SIGNATORY OR SIGNATORIES I certify, in accordance with 37 CFR 1.4(d)(4) that I am: ◉ An attorney or agent registered to practice before the Patent and Trademark Office who is of record in this application Registration Number 71659 ○ A sole inventor ○ A joint inventor; I certify that I am authorized to sign this submission on behalf of all of the inventors as evidenced by the power of attorney in the application ○ A joint inventor; all of whom are signing this request

Signature	/David D. Hsu/
Name	David D. Hsu

*Statement under 37 CFR 3.73(b) is required if terminal disclaimer is signed by the assignee (owner).
Form PTO/SB/96 may be used for making this certification. See MPEP § 324.

Electronic Patent Application Fee Transmittal

Application Number:	17379544
Filing Date:	19-Jul-2021
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Filer:	David Dah-Wei Hsu/Heather Orosco
Attorney Docket Number:	080015-028420US-1249377

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
STATUTORY OR TERMINAL DISCLAIMER	1814	1	170	170
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**170**

Doc Code: DISQ.E.FILE
Document Description: Electronic Terminal Disclaimer – Approved

Application No.: 17379544

Filing Date: 19-Jul-2021

Applicant/Patent under Reexamination: Lo

Electronic Terminal Disclaimer filed on April 27, 2022

☒ APPROVED

This patent is subject to a terminal disclaimer

☐ DISAPPROVED

Approved/Disapproved by: Electronic Terminal Disclaimer automatically approved by EFS-Web

U.S. Patent and Trademark Office

Electronic Acknowledgement Receipt

EFS ID:	45580295
Application Number:	17379544
International Application Number:	
Confirmation Number:	2092
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Customer Number:	20350
Filer:	David Dah-Wei Hsu/Heather Orosco
Filer Authorized By:	David Dah-Wei Hsu
Attorney Docket Number:	080015-028420US-1249377
Receipt Date:	27-APR-2022
Filing Date:	19-JUL-2021
Time Stamp:	19:21:09
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$170
RAM confirmation Number	E20224QJ21043127
Deposit Account	201430
Authorized User	Heather Orosco

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)

37 CFR 1.19 (Document supply fees)

37 CFR 1.17 (Patent application and reexamination processing fees)

37 CFR 1.16 (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	Terminal Disclaimer-Filed (Electronic)	eTerminal-Disclaimer.pdf	41196 311205ba66911b3f65d835b9868e010eaae9ea36	no	2
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	37841 22a858a02ee31d21693a786b29227b1cec3dbb76	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	79037		

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.

I hereby certify that this correspondence is being filed via EFS-Web with the United States Patent and Trademark Office on April 27, 2022.

KILPATRICK TOWNSEND & STOCKTON LLP

By: /Heather A. Orosco/
Heather A. Orosco

PATENT
Attorney Docket No.: 080015-028420US-1249377
Client Ref. No.: 19/MED/901 US Con1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Confirmation No.: 2092
Yuk-Ming Dennis Lo et al.	Examiner: Laura B. Goddard
Application No.: 17/379,544	Technology Center/Art Unit: 1642
Filed: July 19, 2021	
For: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS	**AMENDMENT / RESPONSE TO OFFICE ACTION**
Customer No.: 20350	

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

Commissioner:

In response to the Office Action mailed February 4, 2022, please enter the following amendments and consider the following remarks:

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

Amendments to the Claims are reflected in the listing of claims, which begins on page 3 of this paper.

Amendments to the Drawings begin on page 10 of this paper and include attached Replacement Sheets.

Remarks/Arguments begin on page 11 of this paper.

AMENDMENTS TO THE SPECIFICATION

Please delete paragraph **[0014]**.

[0014] ~~The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.~~

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings of claims in the application.

Listing of Claims:

1. (Previously Presented) A method for determining whether a sample nucleic acid molecule is of fetal or maternal origin, wherein the sample nucleic acid molecule is cell-free and obtained from a biological sample of a female subject pregnant with a fetus, the method comprising:

(a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in the sample nucleic acid molecule and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(b) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:

receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,

apply optimized parameters of the model to the properties of the input data structure; and

output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation;

(d) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, the methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(e) determining whether the sample nucleic acid molecule is of fetal or maternal origin using the methylation statuses of the one or more nucleotides.

2. (Original) The method of claim 1, wherein determining whether the sample nucleic acid molecule is of fetal or maternal origin comprises:

determining a methylation level of the sample nucleic acid molecule using the methylation statuses of the one or more nucleotides; and

comparing the methylation level of the sample nucleic acid molecule to a reference value.

3. (Original) The method of claim 2, wherein the reference value is determined from a methylation level of one or more maternal nucleic acid molecules.

4. (Original) The method of claim 2, wherein:

comparing the methylation level of the sample nucleic acid molecule to the reference value comprises determining the methylation level of the sample nucleic acid molecule is lower than the reference value, and

determining whether the sample nucleic acid molecule is of fetal or maternal origin comprises determining the sample nucleic acid molecule is of fetal origin using the

comparison.

5. (Original) The method of claim 1, further comprising:

identifying the sample nucleic acid molecule as aligning to a predetermined genomic region.

6. (Original) The method of claim 1, wherein:

the sample nucleic acid molecule is one sample nucleic acid molecule of a plurality of sample nucleic acid molecules,

the method further comprising:

performing steps (a)-(e) for each of the plurality of sample nucleic acid molecules to determine whether each of the plurality of sample nucleic acid molecules is of fetal or maternal origin using the methylation statuses, and

determining a fetal fraction using the determination of the fetal or maternal origin of the plurality of sample nucleic acid molecules.

7. (Original) The method of claim 1, wherein the model includes a machine learning model, a principal component analysis, a convolutional neural network, or a logistic regression.

8. (Original) The method of claim 1, wherein the nucleotides within the window are determined using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

9. (Original) The method of claim 1, wherein nucleotides within the window are determined without using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

10. (Original) The method of claim 1, wherein the optical signal is a fluorescence signal from a dye-labeled nucleotide.

11. (Original) The method of claim 1, wherein each window associated with each input data structure comprises at least 4 consecutive nucleotides on a first strand of each

first nucleic acid molecule.

12. (Original) The method of claim 1, wherein the methylation is 5mC or 6mA.

13. (Original) The method of claim 1, wherein:

determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, and

the sample nucleic acid molecule is a first nucleic acid molecule;

the method further comprising:

providing a second nucleic acid molecule, the second nucleic acid molecule being identical to the first nucleic acid molecule and having a methylated nucleotide at a position corresponding to the target position,

performing bisulfite sequencing on the second nucleic acid molecule to obtain bisulfite sequencing results, and

verifying the methylation present in the nucleotide at the target position using bisulfite sequencing results.

14-17. (Canceled)

18. (Previously Presented) A method for determining whether a sample nucleic acid molecule is of fetal or maternal origin, wherein the sample nucleic acid molecule is cell-free and obtained from a biological sample of a female subject pregnant with a fetus, the method comprising:

(a) sequencing the sample nucleic acid molecule by measuring pulses in an optical signal corresponding to nucleotides in the sample nucleic acid molecule;

(b) receiving data acquired by sequencing the sample nucleic acid molecule, and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid

molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(d) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:

receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,

apply optimized parameters of the model to the properties of the input data structure; and

output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation;

(e) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, the methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(f) determining whether the sample nucleic acid molecule is of fetal or maternal origin using the methylation statuses of the one or more nucleotides.

19. (Original) The method of claim 18, further comprising:

cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and

ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule to form a circular DNA molecule, wherein:

the circular DNA molecule is the sample nucleic acid molecule,

the window of the nucleotides corresponding to the input data structure comprises nucleotides on a first strand of the sample nucleic acid molecule and nucleotides on a second strand of the sample nucleic acid molecule, and

the input data structure further comprises for each nucleotide within the window a value of a strand property, the strand property indicating the nucleotide being present on either the first strand or the second strand.

20. (Original) The method of claim 18, further comprising enriching or filtering each nucleotide within the window.

21. (Original) The method of claim 20, comprising enriching each nucleotide within the window by cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule.

22. (Original) The method of claim 20, comprising filtering each nucleotide within the window selecting double-stranded DNA molecules having a size within a size range.

23. (Original) The method of claim 18, wherein sequencing the sample nucleic acid molecule comprises:

providing a polymerase that is capable of processing a first nucleic acid molecule, wherein the first nucleic acid molecule comprises the methylation in at least one nucleotide;

incorporating dye-labeled nucleotides into the first nucleic acid molecule to form the sample nucleic acid molecule; and

measuring pulses in the optical signal corresponding to the dye-labeled nucleotides,

wherein determining whether the methylation is present in the nucleotide at the

target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, thereby determining the methylation is present in the at least one nucleotide in the first nucleic acid molecule.

24. (Original) The method of claim 23, wherein the polymerase is immobilized.

25. (Original) The method of claim 23, further comprising ligating Watson and Crick strands of the first nucleic acid molecule to form a circular template molecule comprising a double-stranded segment having each end capped with a hairpin adaptor, wherein the circular template molecule is the first nucleic acid molecule.

26. (Original) The method of claim 18, further comprising:

providing a double-stranded nucleic acid molecule comprising the methylation in at least one nucleotide; and

ligating Watson and Crick strands of the double-stranded nucleic acid molecule to form a circular template molecule comprising a double-stranded segment having each end capped with a hairpin adaptor, wherein the circular template molecule is the sample nucleic acid molecule; wherein:

the data comprises sequences of the Watson and Crick strands of the double-stranded nucleic acid molecule contained in the circular template molecule, and

determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, thereby determining the methylation is present in the at least one nucleotide in the double-stranded nucleic acid molecule.

27. (Previously Presented) The method of claim 18, further comprising aligning the sequenced nucleotides to a reference genome to provide the positions of the nucleotides within the sample nucleic acid molecule.

28-30. (Canceled)

Electronic Acknowledgement Receipt

EFS ID:	45580159
Application Number:	17379544
International Application Number:	
Confirmation Number:	2092
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Customer Number:	20350
Filer:	David Dah-Wei Hsu/Heather Orosco
Filer Authorized By:	David Dah-Wei Hsu
Attorney Docket Number:	080015-028420US-1249377
Receipt Date:	27-APR-2022
Filing Date:	19-JUL-2021
Time Stamp:	19:15:55
Application Type:	Utility under 35 USC 111(a)

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		1249377_Response_to_Non-Final_Office_Action.pdf	105813 babfc462a8c6413f004c87a55386c46db882a2ca	yes	12

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Amendment/Request for Reconsideration-After Non-Final Rejection	1	1
	Specification	2	2
	Claims	3	9
	Drawings-only black and white line drawings	10	10
	Applicant Arguments/Remarks Made in an Amendment	11	12

Warnings:					
Information:					
2	Drawings-only black and white line drawings	1249377_Replacement_Sheets_part_1_of_2.pdf	4030910 515fa5bedf8d61a49e6da2c608d73aa17f82adab	no	43
Warnings:					
Information:					
3	Drawings-only black and white line drawings	1249377_Replacement_Sheets_part_2_of_2.pdf	3411088 71b192ca5851461a50271233d4759f844c04d2e7	no	43
Warnings:					
Information:					
		Total Files Size (in bytes):	7547811		

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.

REMARKS/ARGUMENTS

Upon entry of the present amendment, claims 1-13 and 18-27 will be pending in this application. Claims 14-17 and 28-30 have been canceled. No new matter is added. Based on the following remarks, Applicant respectfully requests reconsideration and allowance of the pending claims.

Objection to Color Drawings/Specification

The color petition of July 19, 2021 was not granted, and the Office therefore objects to the color drawings. Substitute black and white drawings are provided. No new matter has been added. Withdrawal of this objection is respectfully requested.

Double Patenting

Claims 1-13 and 18-27 stand rejected on the ground of nonstatutory double patenting over claims 1-19 of U.S. Patent No. 11,091,794. Applicant is hereby filing a terminal disclaimer over U.S. Patent No. 11,091,794. For at least these reasons, Applicant requests withdrawal of the double patenting rejection.

CONCLUSION

In view of the foregoing, Applicant believes all claims now pending in this application are in condition for allowance. The issuance of a formal Notice of Allowance at an early date is respectfully requested.

Except for the issue fees payable under 37 C.F.R. § 1.18, the Director is authorized to charge any additional fees during pendency of this application, including any required extension of time fees, or credit any overpayment to Deposit Account Number 20-1430. This paragraph is intended to be a constructive petition for extension of time in accordance with 37 C.F.R. § 1.136(a)(3).

If the Examiner believes a telephone conference would expedite prosecution of this application, please contact the undersigned at (303) 571-4000 or dhsu@kilpatricktownsend.com.

Respectfully submitted,

/David D. Hsu/
David D. Hsu
Registration No. 71659

KILPATRICK TOWNSEND & STOCKTON LLP

DDH:hao

75724415V.1

AMENDMENT(S) TO THE DRAWINGS

The attached sheets of drawings include changes to FIGS. 1-16, 18A-26, 28-40, 42, 44, 47, 49-50, 52A, 54, 56A-58, 61, 63A-64B, 67A-B, 70A-B, 73-77, 79A-B, 79D, 80D, 82-85C, 88-94, 104, 108, 114-115, 117, 126, 130, and 131. These sheets replace the original sheets.

Attachment(s): Replacement Sheets

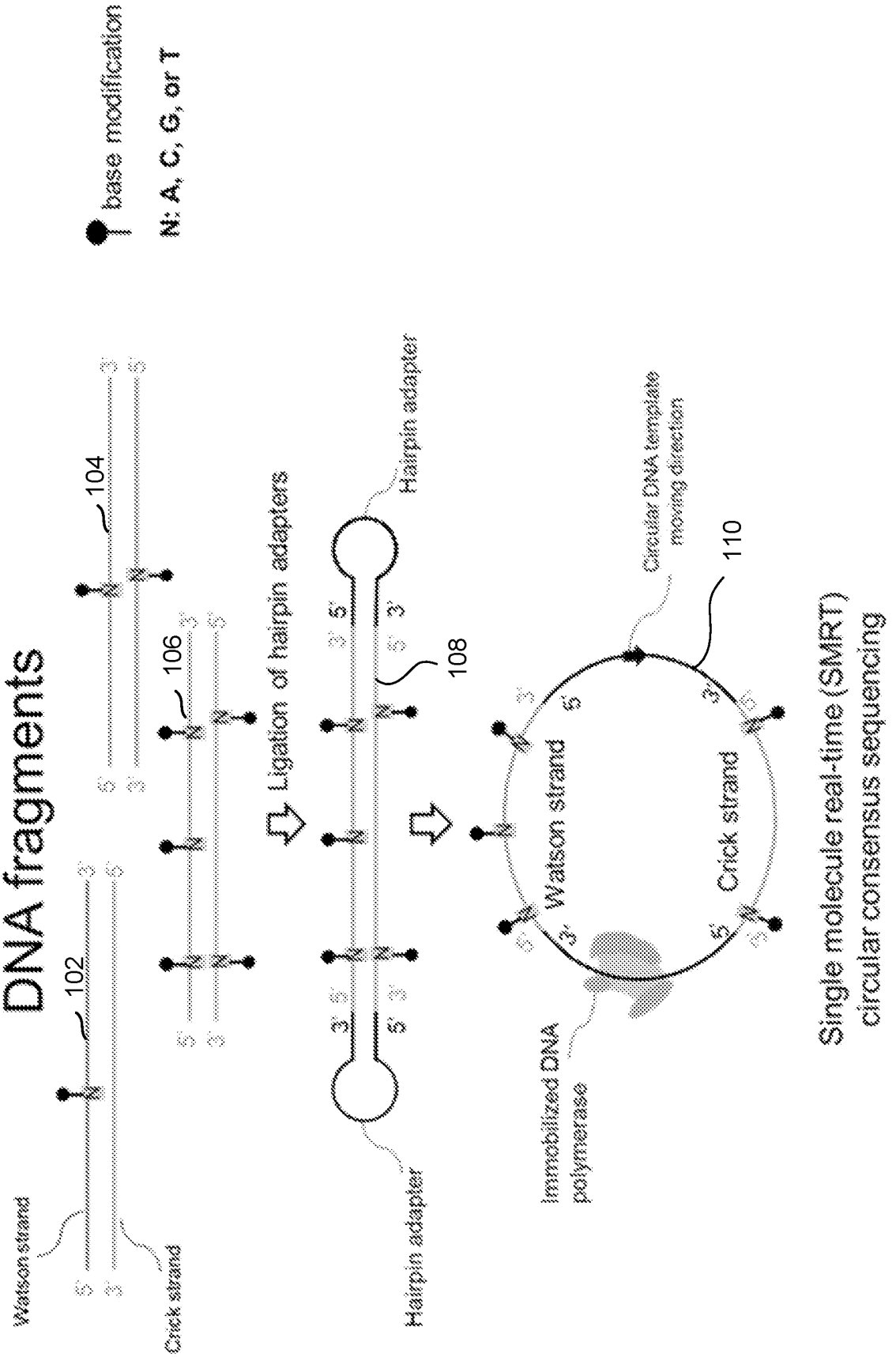

base modification
N: A, C, G, or T
DNA fragments
102
104
106
Watson strand
Crick strand
Ligation of hairpin adapters
Hairpin adapter
Hairpin adapter
108
Immobilized DNA polymerase
Watson strand
Crick strand
Circular DNA template moving direction
110
Single molecule real-time (SMRT) circular consensus sequencing

FIG. 1

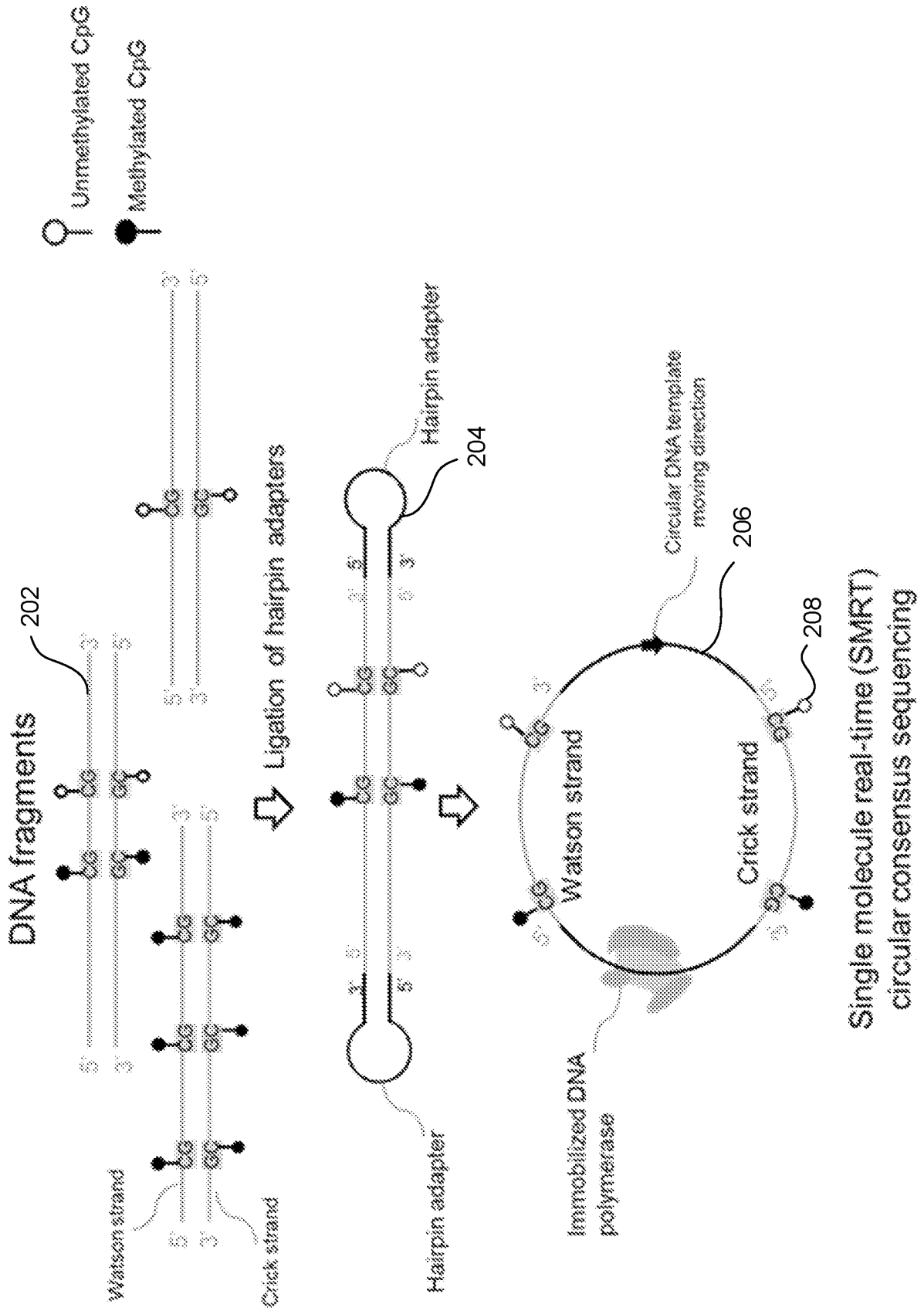

Unmethylated CpG
Methylated CpG
DNA fragments
202
Watson strand
Crick strand
Ligation of hairpin adapters
Hairpin adapter
204
Hairpin adapter
Immobilized DNA polymerase
Watson strand
Crick strand
Circular DNA template moving direction
206
208
Single molecule real-time (SMRT) circular consensus sequencing

FIG. 2

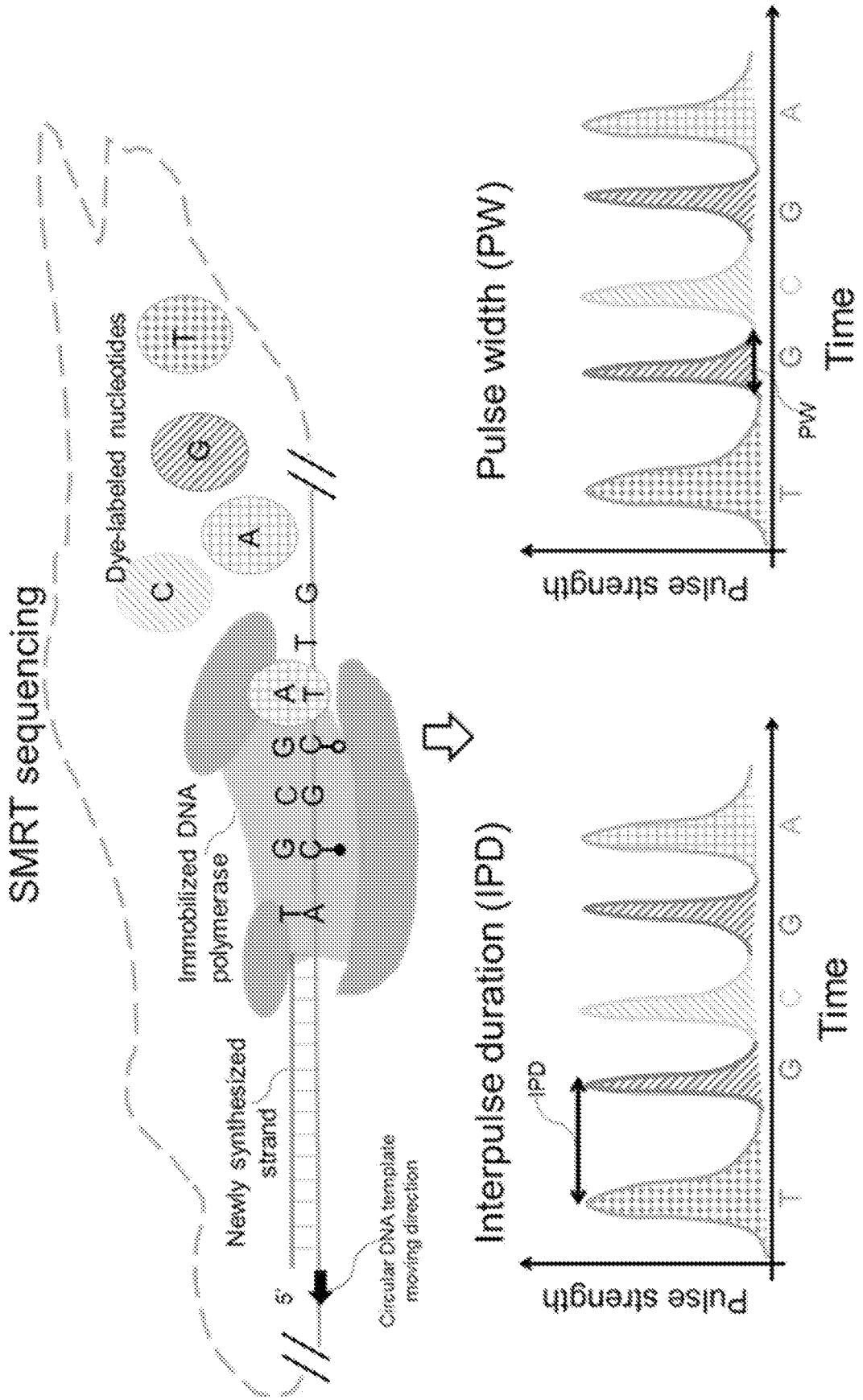

SMRT sequencing
Dye-labeled nucleotides
C
A
G
T
Immobilized DNA polymerase
Newly synthesized strand
5'
Circular DNA template moving direction
Pulse width (PW)
Pulse strength
Time
PW
Interpulse duration (IPD)
IPD

FIG. 3

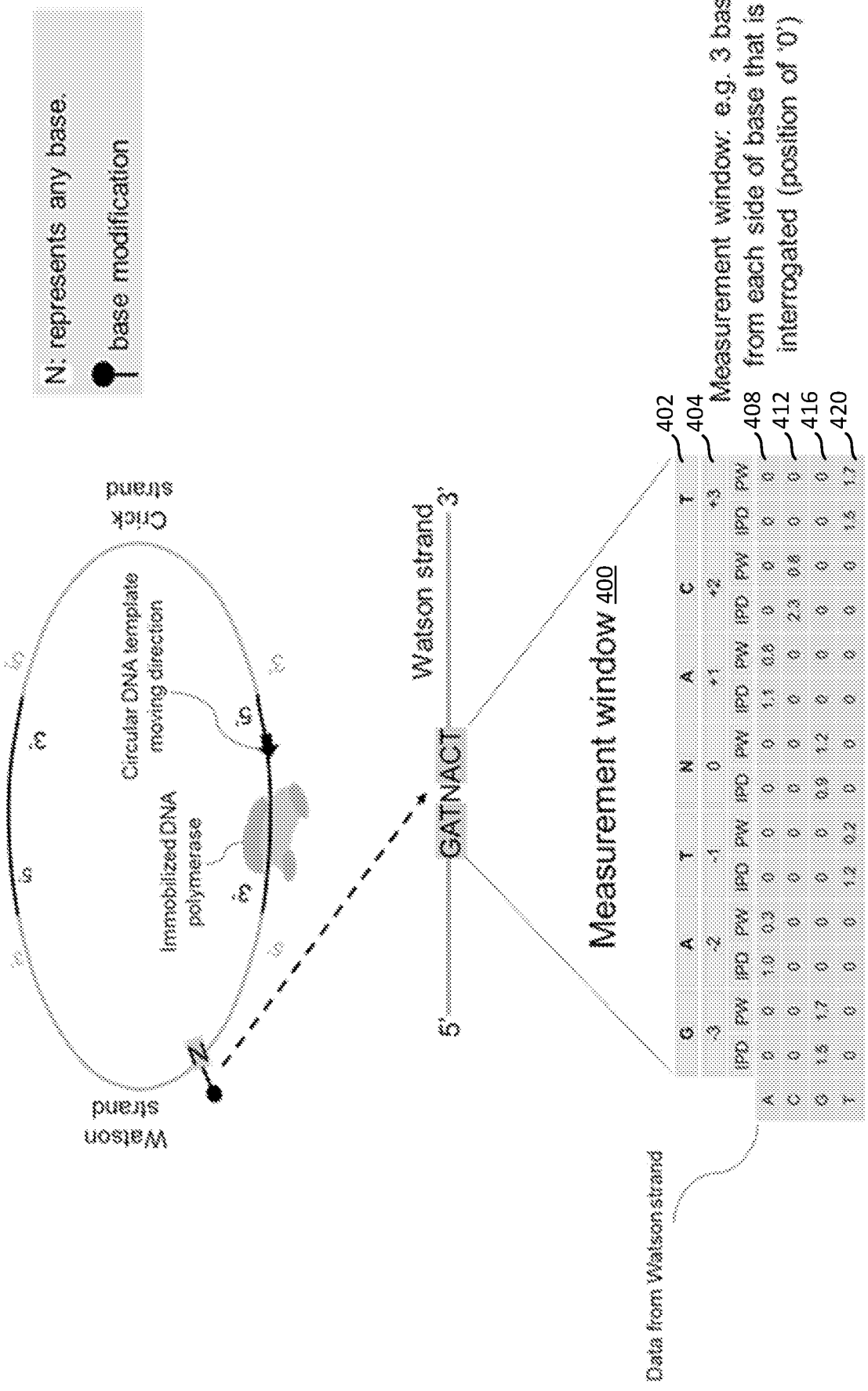

N: represents any base.
base modification
Watson strand
Nick strand
Circular DNA template moving direction
Immobilized DNA polymerase
Watson strand
5'
3'
GATNACT
Measurement window 400
Measurement window: e.g. 3 bases from each side of base that is interrogated (position of '0')
402
404
408
412
416
420
Data from Watson strand

FIG. 4

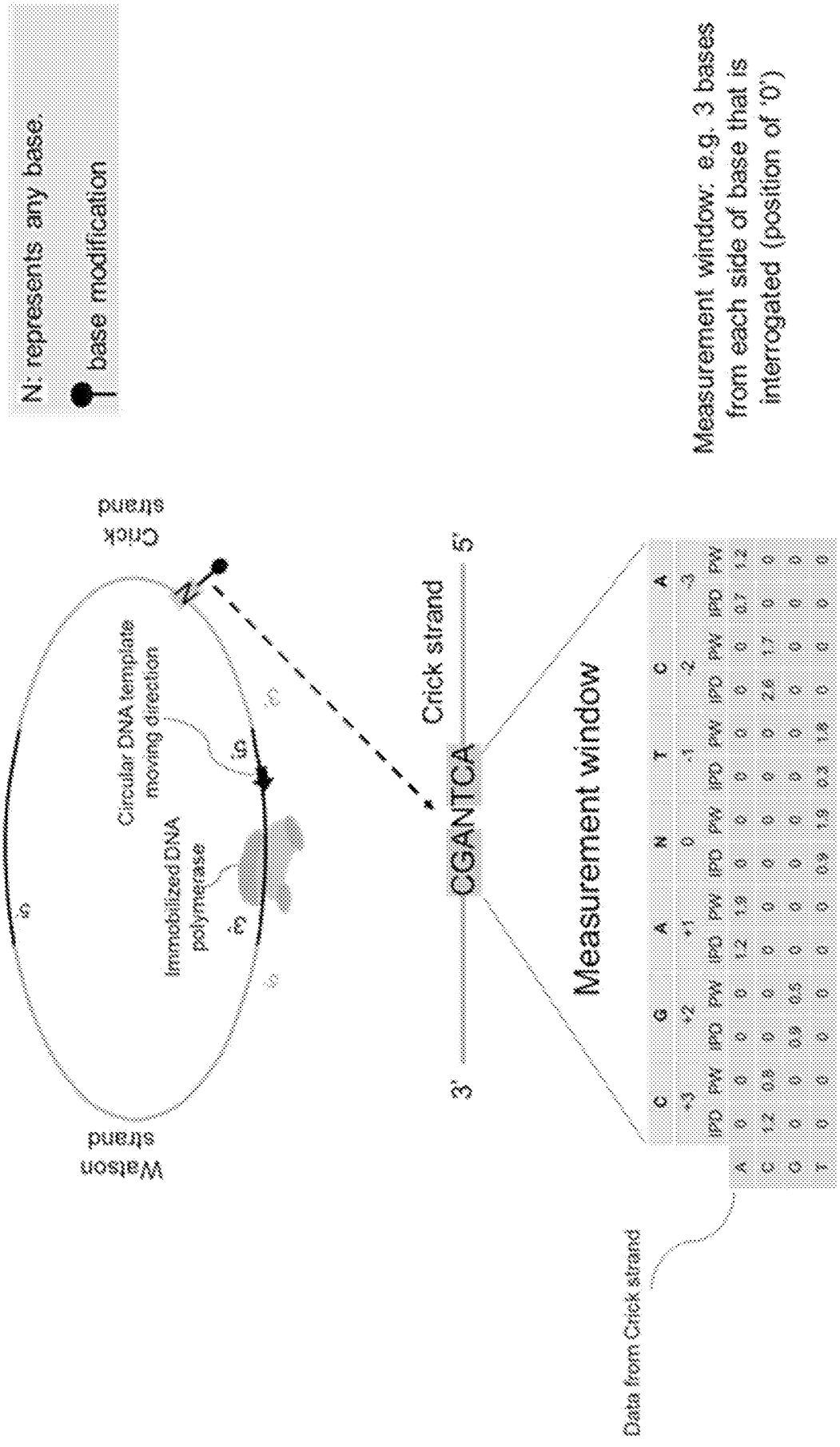

N: represents any base.
base modification
Measurement window: e.g. 3 bases from each side of base that is interrogated (position of '0')
Nick strand
Circular DNA template moving direction
Immobilized DNA polymerase
Watson strand
Crick strand
5'
3'
CGANTCA
Measurement window
Data from Crick strand

FIG. 5

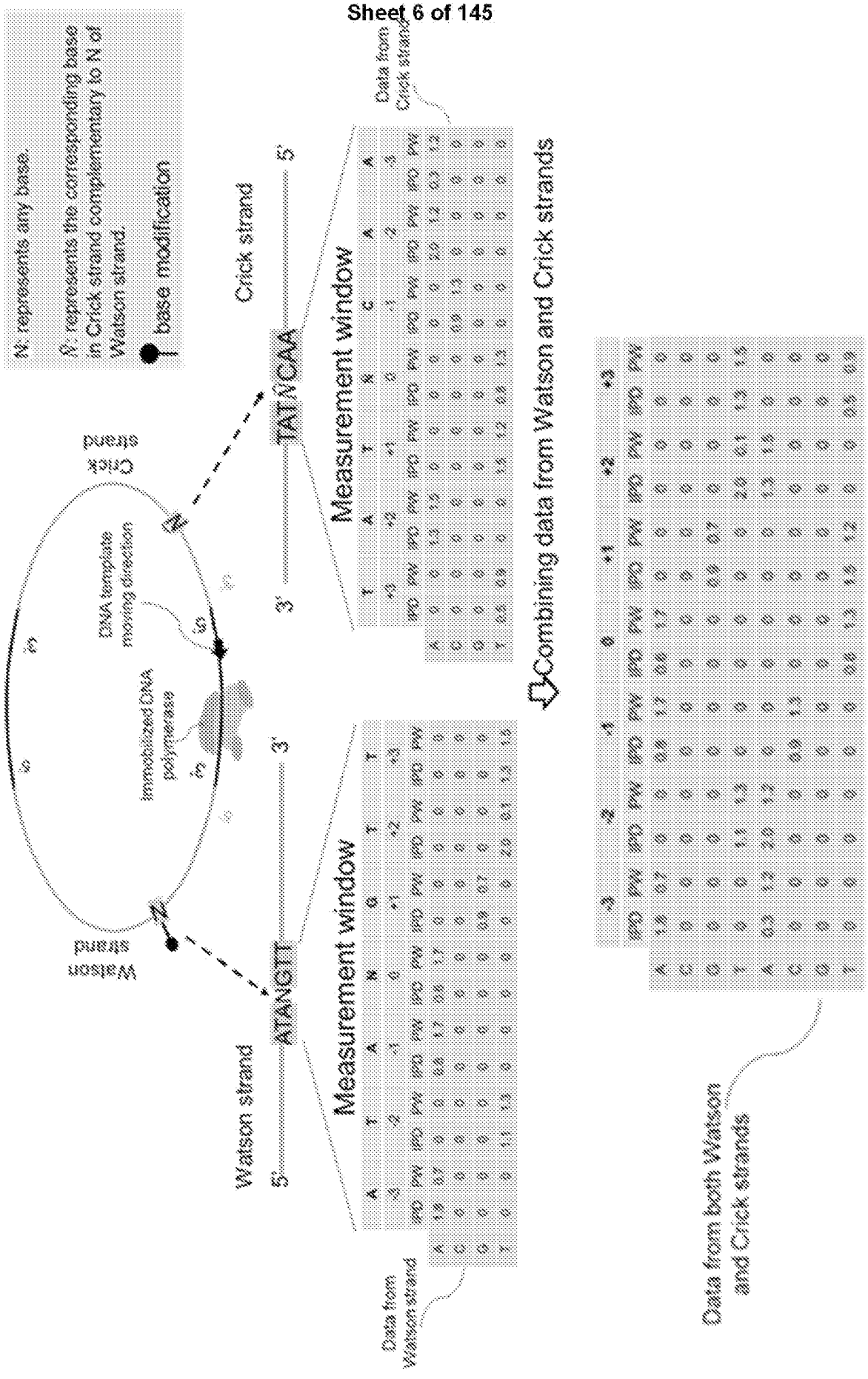

N: represents any base.
δ: represents the corresponding base in Crick strand complementary to N of Watson strand.
base modification
Watson strand
Crick strand
DNA template moving direction
Immobilized DNA polymerase
5'
3'
ATANGTT
TATNCAA
Measurement window
Data from Watson strand
Data from Crick strand
Combining data from Watson and Crick strands
Data from both Watson and Crick strands

FIG. 6

N: represents any base
N': represents a base proximal to the corresponding base in Crick strand complementary to N of Watson strand
base modification
Watson strand
Crick strand
Circular DNA template moving direction
Immobilized DNA polymerase
5'
3'
ATANGTT
CGCN'ATG
Measurement window
Data from Watson strand
Data from Crick strand
Combining data from Watson and Crick strand
Data from both Watson and Crick strands

FIG. 7

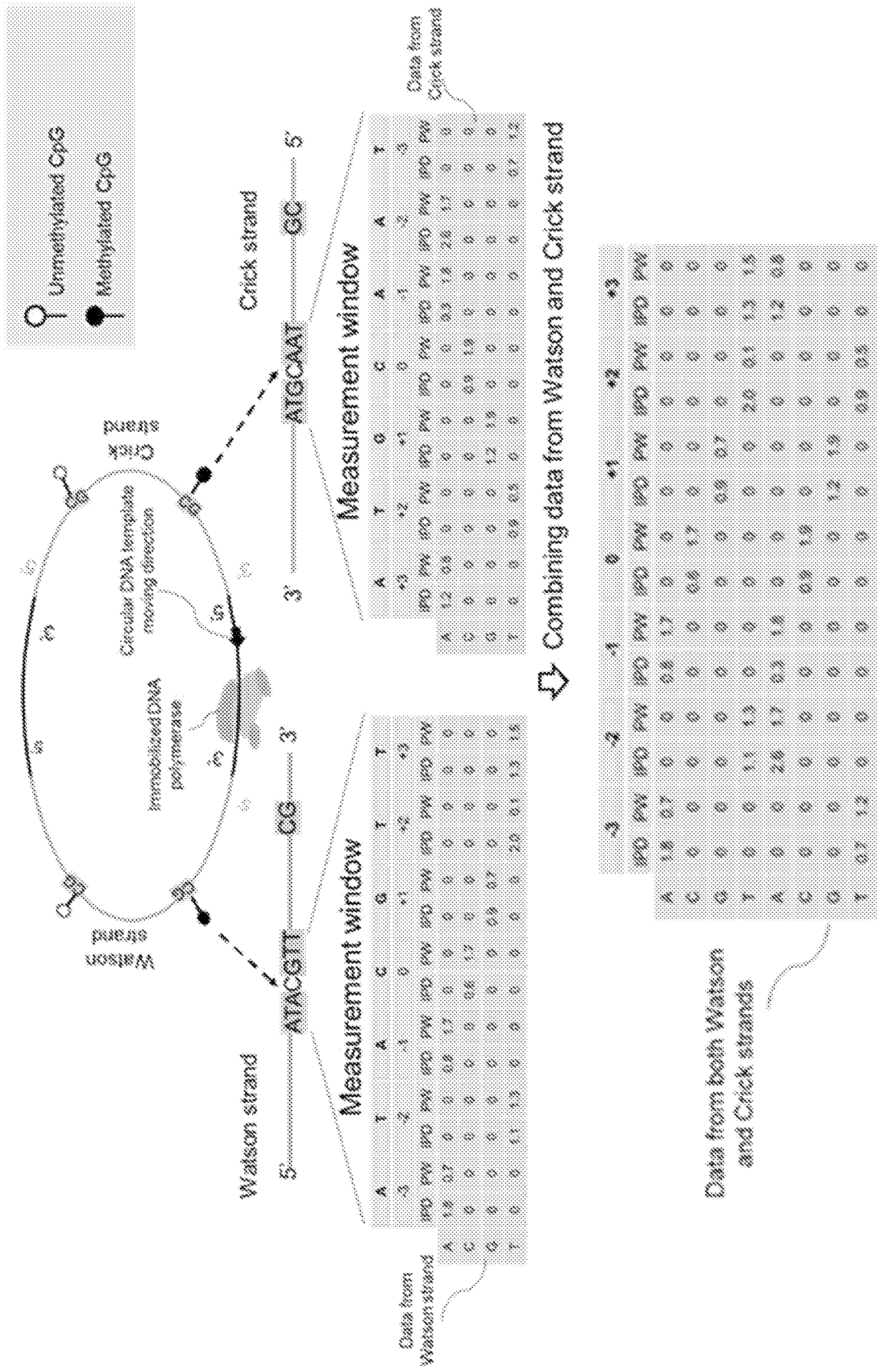

Unmethylated CpG
Methylated CpG
Watson strand
Crick strand
Circular DNA template moving direction
Immobilized DNA polymerase
Watson strand
Crick strand
5'
3'
ATACGTT
CG
ATGCAAT
GC
Measurement window
Measurement window
Data from Watson strand
Data from Crick strand
Combining data from Watson and Crick strand
Data from both Watson and Crick strands

FIG. 8

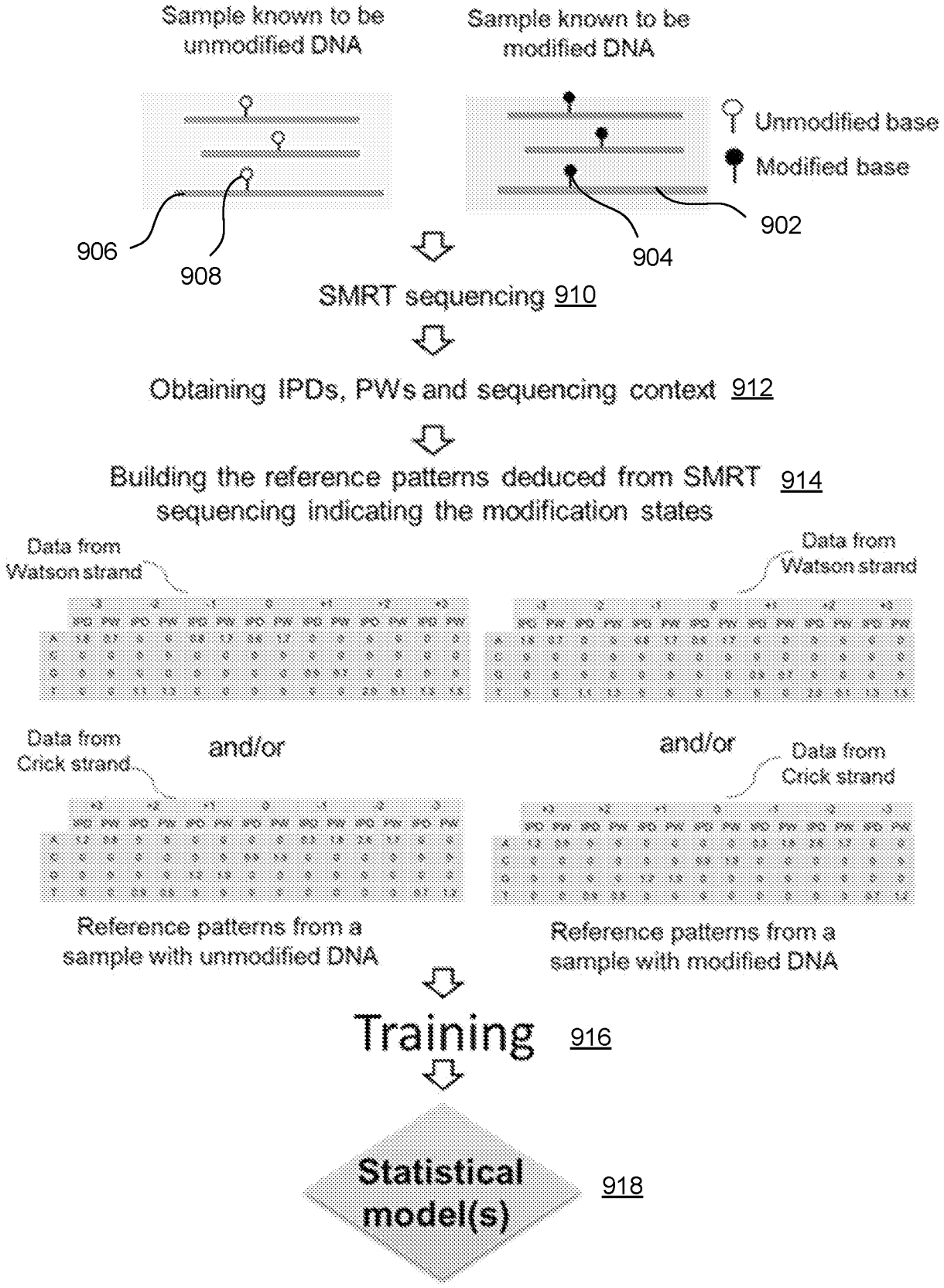

Sample known to be unmodified DNA
Sample known to be modified DNA
Unmodified base
Modified base
906
908
902
904
SMRT sequencing 910
Obtaining IPDs, PWs and sequencing context 912
Building the reference patterns deduced from SMRT sequencing indicating the modification states 914
Data from Watson strand
Data from Watson strand
Data from Crick strand
and/or
and/or
Data from Crick strand
Reference patterns from a sample with unmodified DNA
Reference patterns from a sample with modified DNA
Training 916
Statistical model(s)
918

FIG. 9

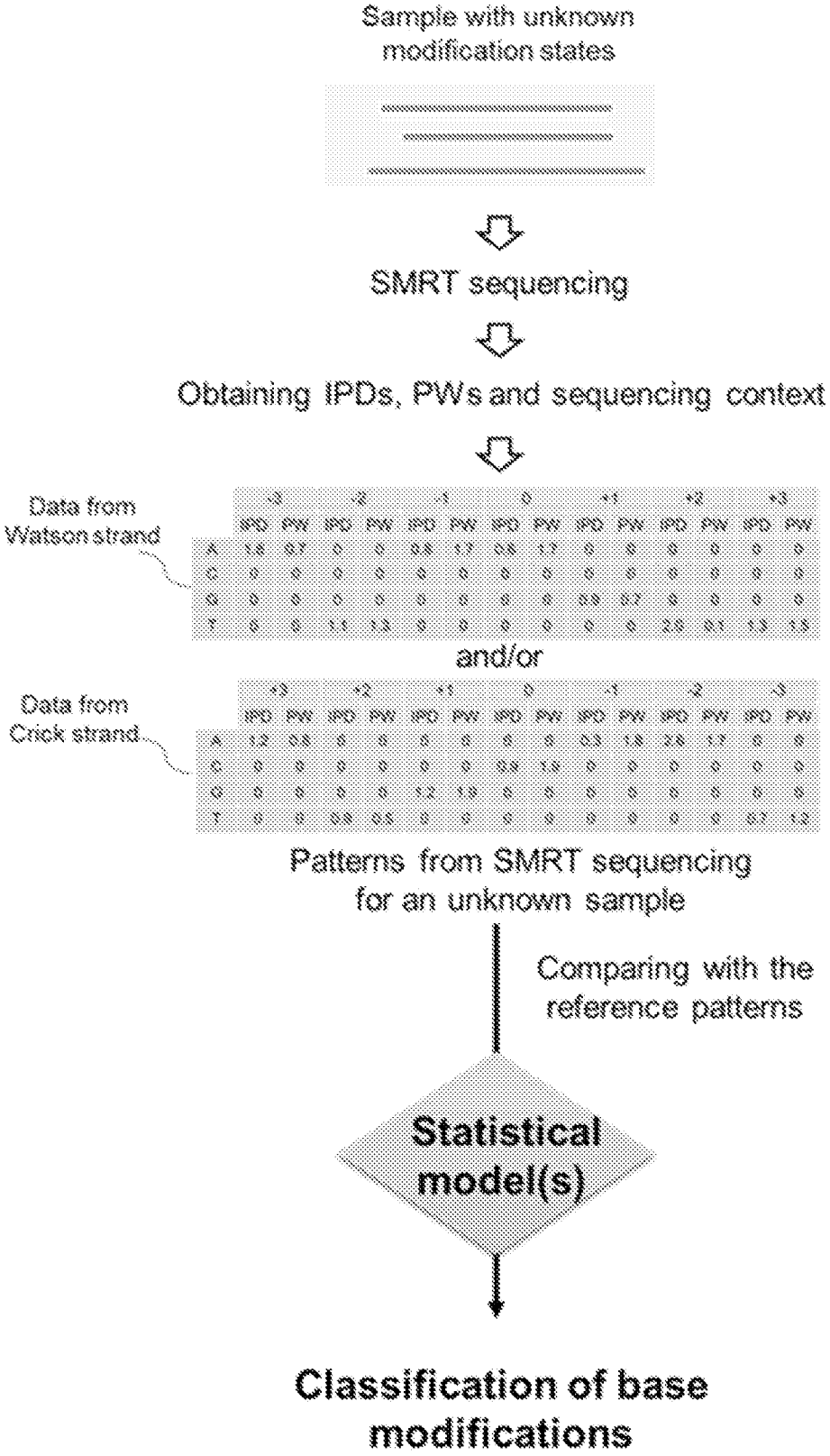

Sample with unknown modification states
SMRT sequencing
Obtaining IPDs, PWs and sequencing context
Data from Watson strand
-3 -2 -1 0 +1 +2 +3
IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW
A 1.6 0.7 0 0 0.5 1.7 0.8 1.7 0 0 0 0 0 0
C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
G 0 0 0 0 0 0 0 0 0.9 0.7 0 0 0 0
T 0 0 1.1 1.3 0 0 0 0 0 0 2.8 0.1 1.3 1.3
and/or
Data from Crick strand
+3 +2 +1 0 -1 -2 -3
IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW
A 1.2 0.8 0 0 0 0 0 0 0.3 1.8 2.8 1.7 0 0
C 0 0 0 0 0 0 0.9 1.9 0 0 0 0 0 0
G 0 0 0 0 1.2 1.8 0 0 0 0 0 0 0 0
T 0 0 0.8 0.5 0 0 0 0 0 0 0 0 0.7 1.2
Patterns from SMRT sequencing for an unknown sample
Comparing with the reference patterns
Statistical model(s)
Classification of base modifications

FIG. 10

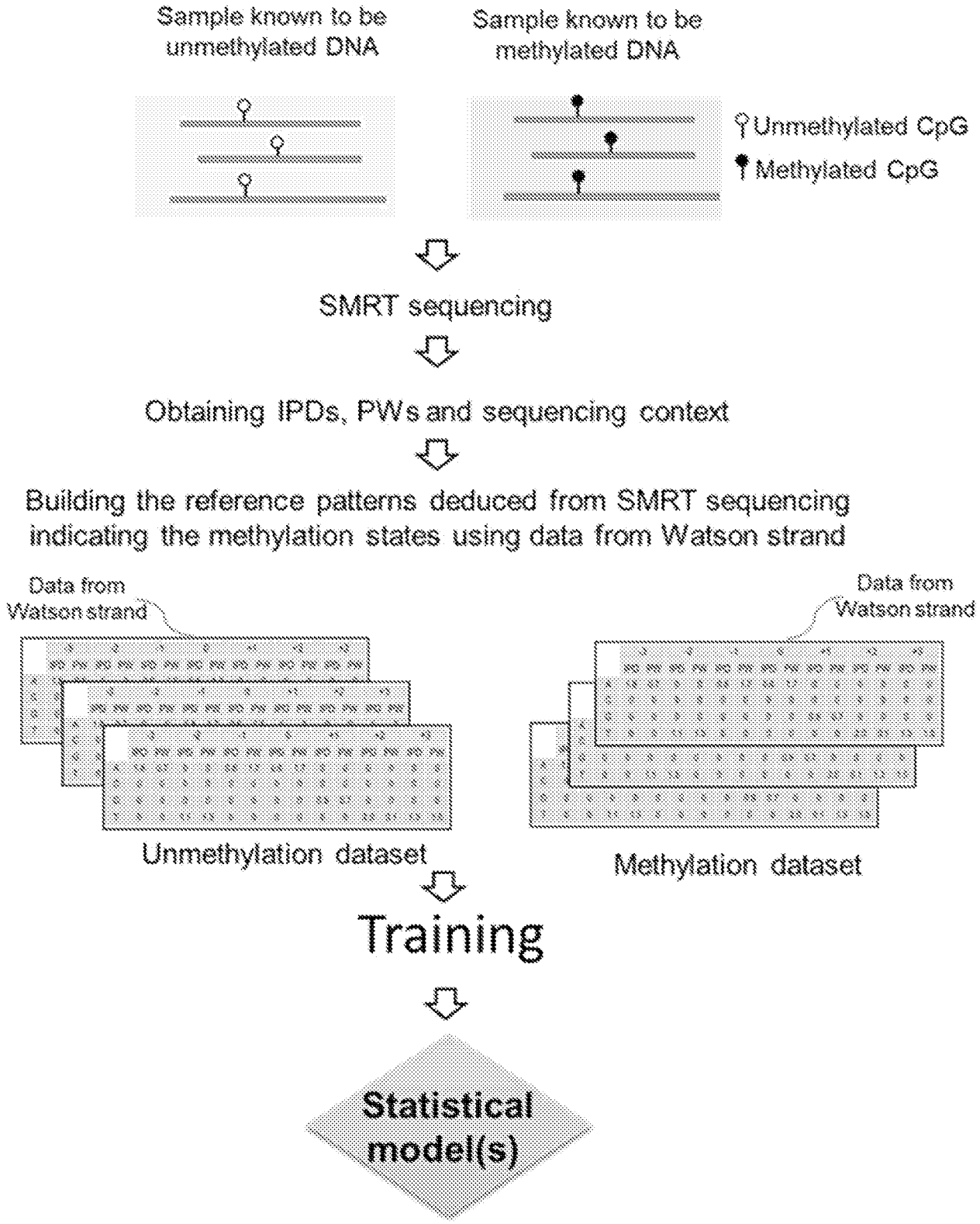
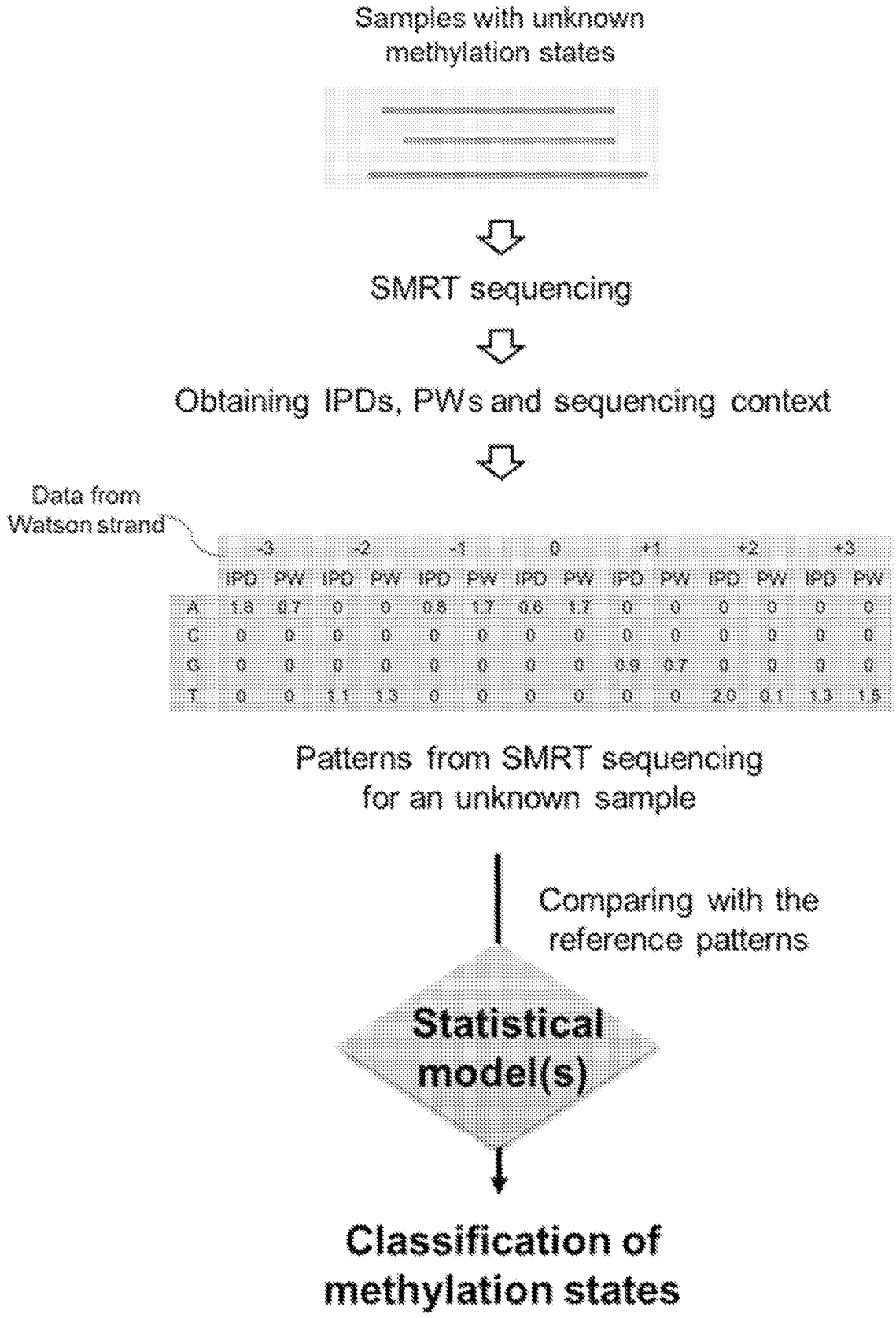

Sample known to be unmethylated DNA
Sample known to be methylated DNA
Unmethylated CpG
Methylated CpG
SMRT sequencing
Obtaining IPDs, PWs and sequencing context
Building the reference patterns deduced from SMRT sequencing indicating the methylation states using data from Watson strand
Data from Watson strand
Data from Watson strand
Unmethylation dataset
Methylation dataset
Training
Statistical model(s)

FIG. 11



	-3		-2		-1		0		+1		+2		+3	
	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW	IPD	PW
A	1.6	0.7	0	0	0.8	1.7	0.6	1.7	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0.9	0.7	0	0	0	0
T	0	0	1.1	1.3	0	0	0	0	0	0	2.0	0.1	1.3	1.5

FIG. 12

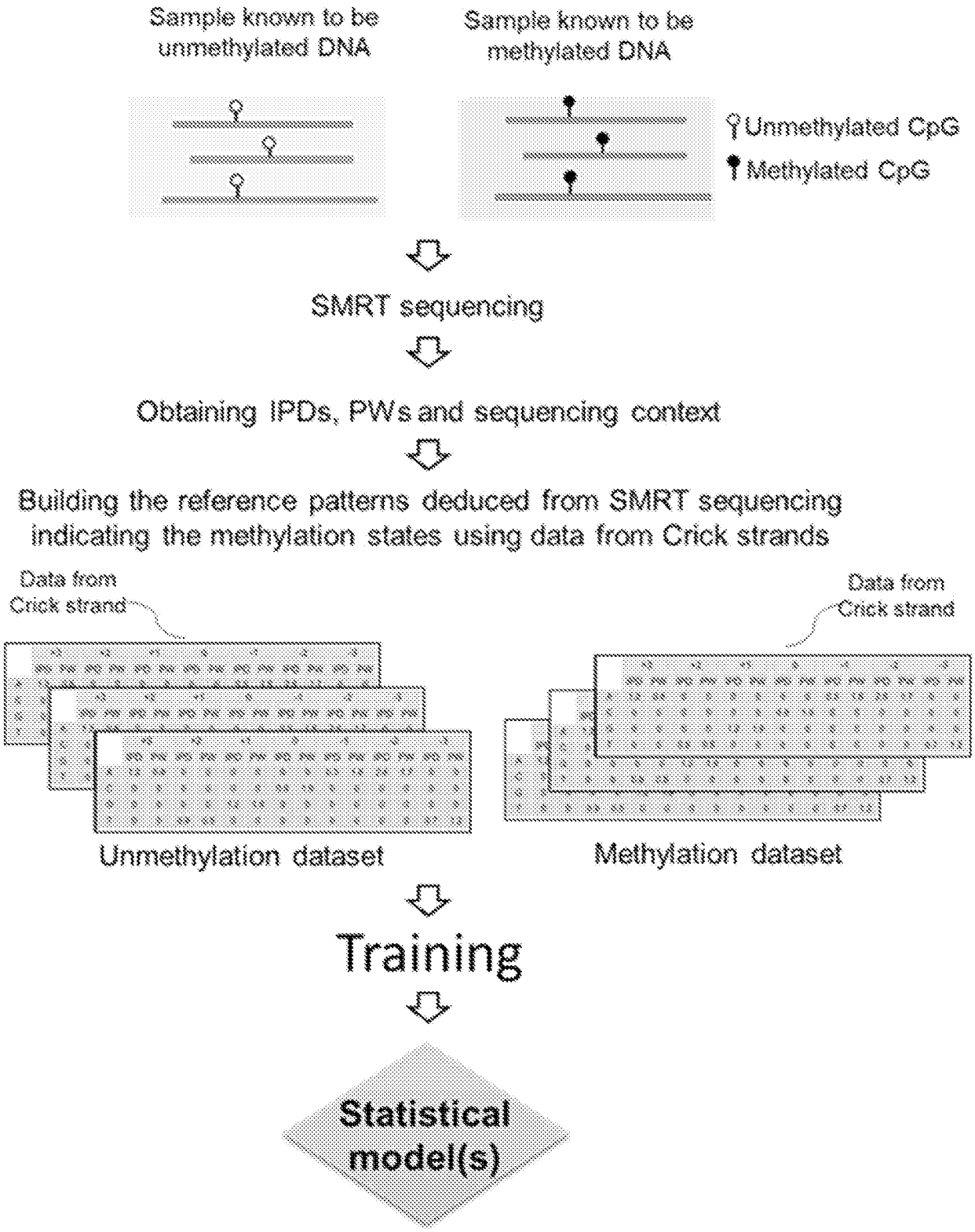

Sample known to be unmethylated DNA
Sample known to be methylated DNA
Unmethylated CpG
Methylated CpG
SMRT sequencing
Obtaining IPDs, PWs and sequencing context
Building the reference patterns deduced from SMRT sequencing indicating the methylation states using data from Crick strands
Data from Crick strand
Data from Crick strand
Unmethylation dataset
Methylation dataset
Training
Statistical model(s)

FIG. 13

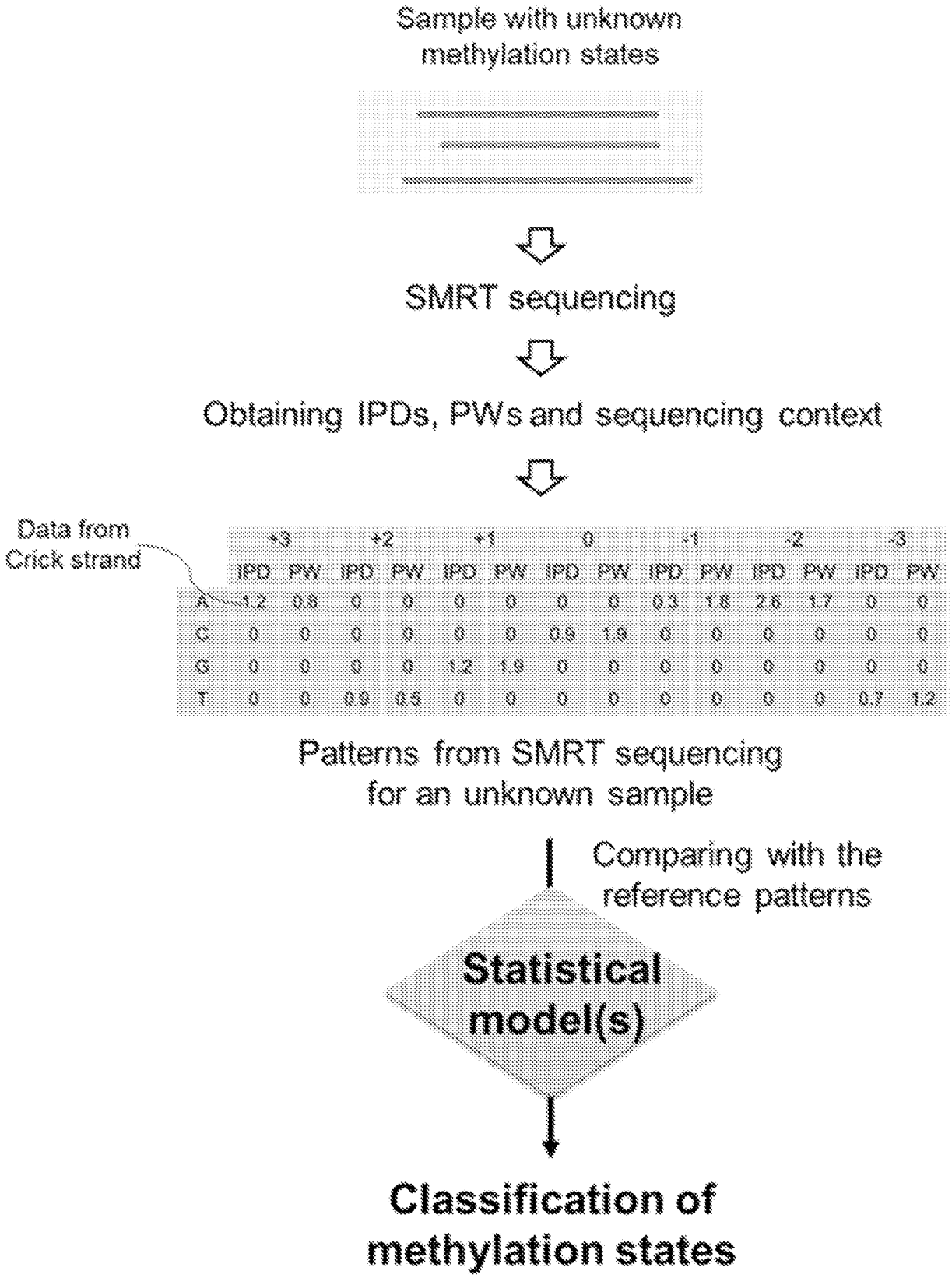

Sample with unknown methylation states
SMRT sequencing
Obtaining IPDs, PWs and sequencing context
Data from Crick strand
+3 +2 +1 0 -1 -2 -3
IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW
A 1.2 0.6 0 0 0 0 0 0 0.3 1.6 2.6 1.7 0 0
C 0 0 0 0 0 0 0.9 1.9 0 0 0 0 0 0
G 0 0 0 0 1.2 1.9 0 0 0 0 0 0 0 0
T 0 0 0.9 0.5 0 0 0 0 0 0 0 0 0.7 1.2
Patterns from SMRT sequencing for an unknown sample
Comparing with the reference patterns
Statistical model(s)
Classification of methylation states

FIG.14

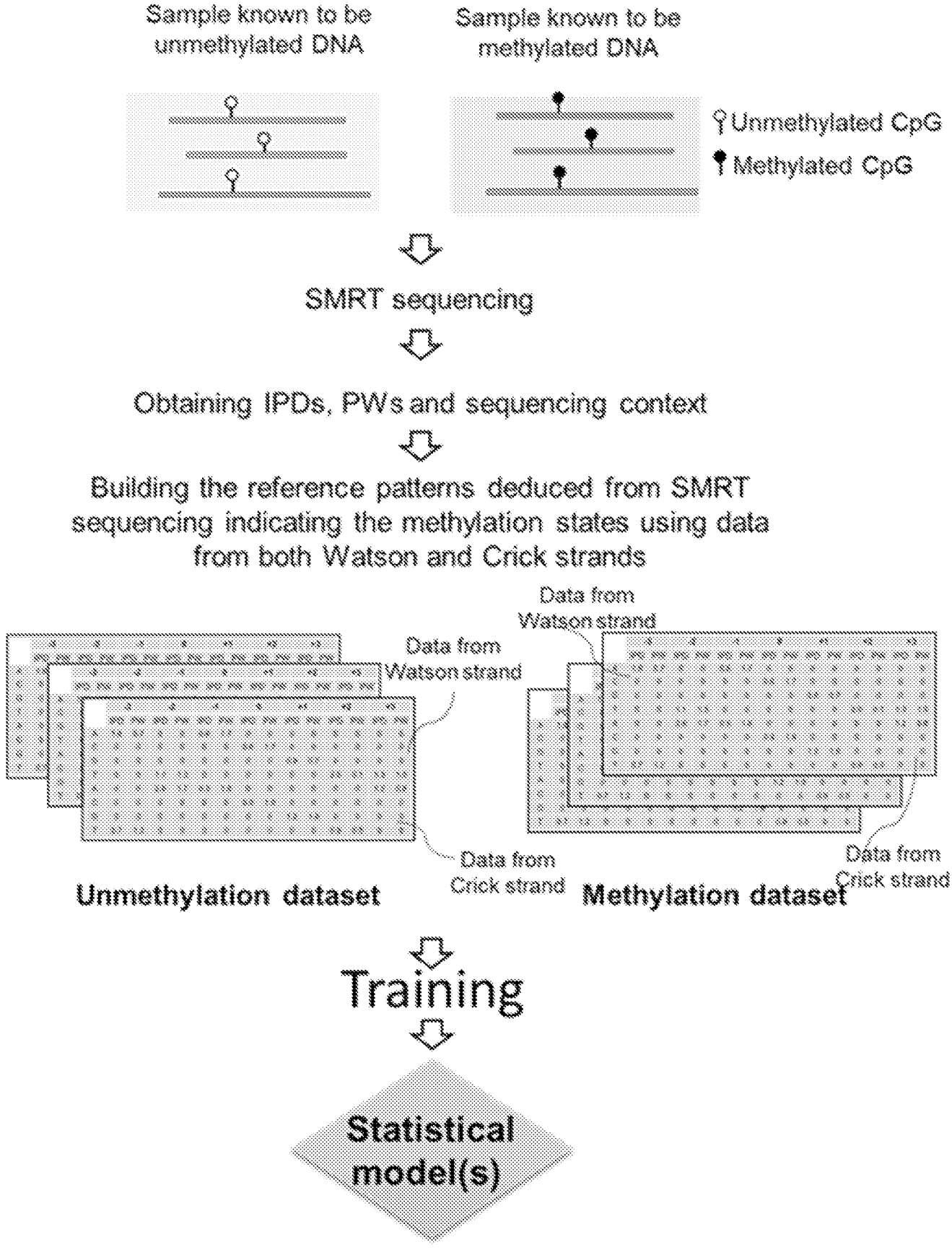

Sample known to be unmethylated DNA
Sample known to be methylated DNA
Unmethylated CpG
Methylated CpG
SMRT sequencing
Obtaining IPDs, PWs and sequencing context
Building the reference patterns deduced from SMRT sequencing indicating the methylation states using data from both Watson and Crick strands
Data from Watson strand
Data from Watson strand
Data from Crick strand
Data from Crick strand
Unmethylation dataset
Methylation dataset
Training
Statistical model(s)

FIG. 15

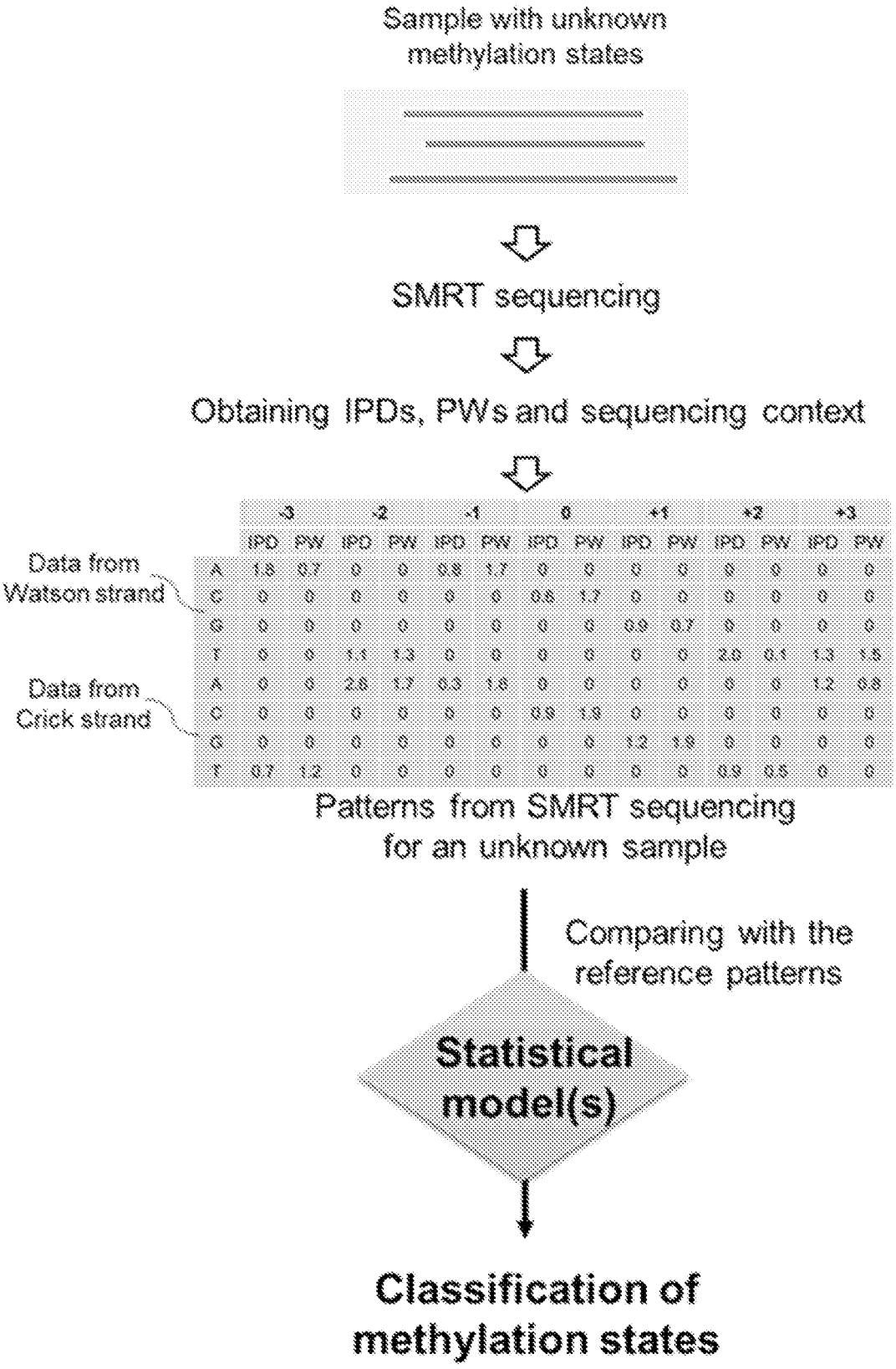

Sample with unknown methylation states
SMRT sequencing
Obtaining IPDs, PWs and sequencing context
Data from Watson strand
Data from Crick strand
-3 -2 -1 0 +1 +2 +3
IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW IPD PW
A 1.8 0.7 0 0 0.8 1.7 0 0 0 0 0 0 0 0
C 0 0 0 0 0 0 0.6 1.7 0 0 0 0 0 0
G 0 0 0 0 0 0 0 0 0.9 0.7 0 0 0 0
T 0 0 1.1 1.3 0 0 0 0 0 0 2.0 0.1 1.3 1.5
A 0 0 2.6 1.7 0.3 1.6 0 0 0 0 0 0 1.2 0.8
C 0 0 0 0 0 0 0.9 1.9 0 0 0 0 0 0
G 0 0 0 0 0 0 0 0 1.2 1.9 0 0 0 0
T 0.7 1.2 0 0 0 0 0 0 0 0 0.9 0.5 0 0
Patterns from SMRT sequencing for an unknown sample
Comparing with the reference patterns
Statistical model(s)
Classification of methylation states

FIG. 16

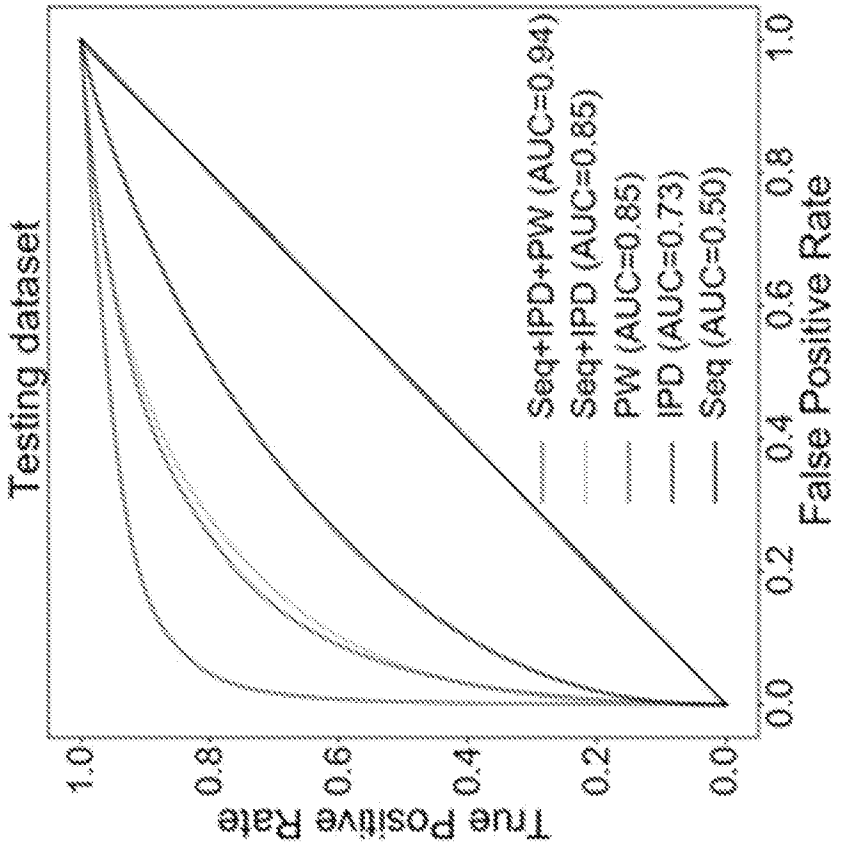


FIG. 18B

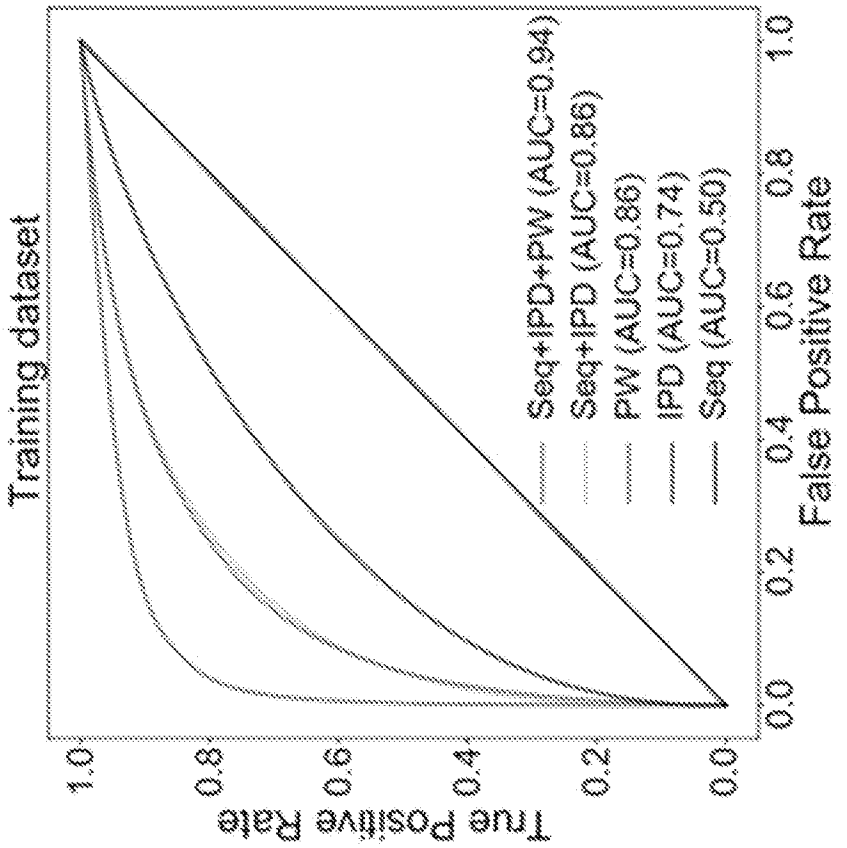


FIG. 18A

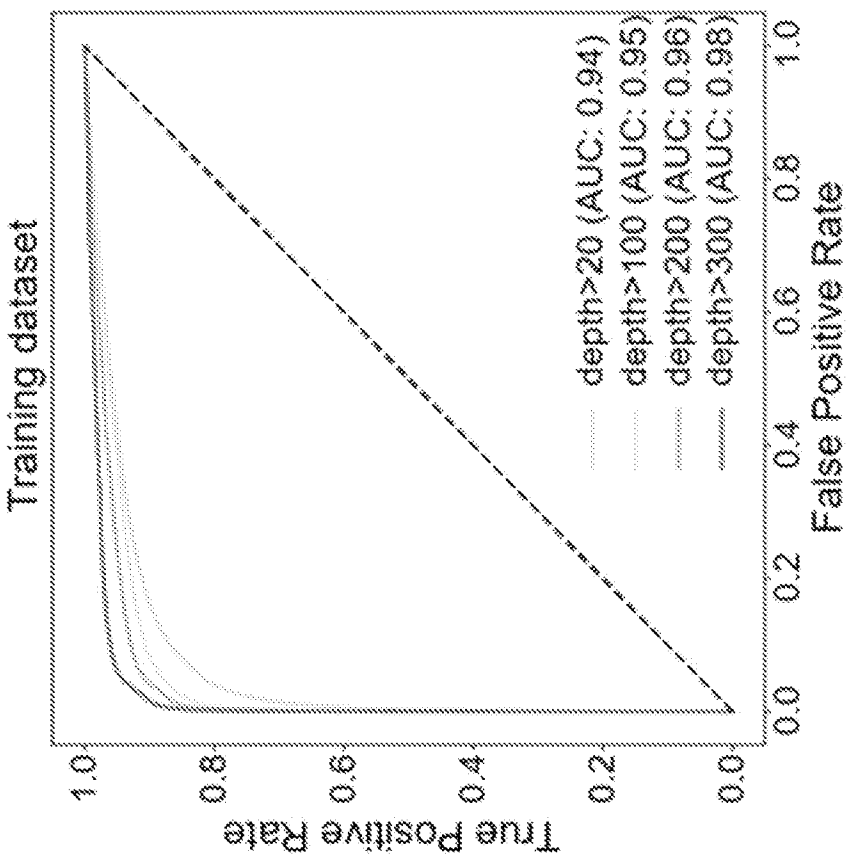


FIG. 19A

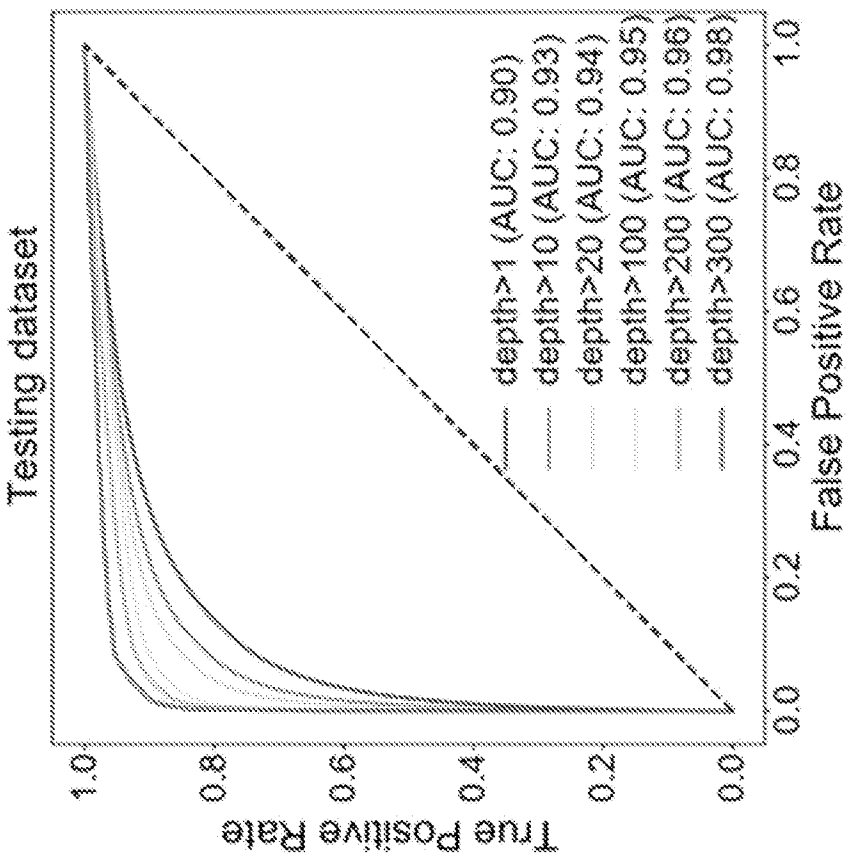


FIG. 19B

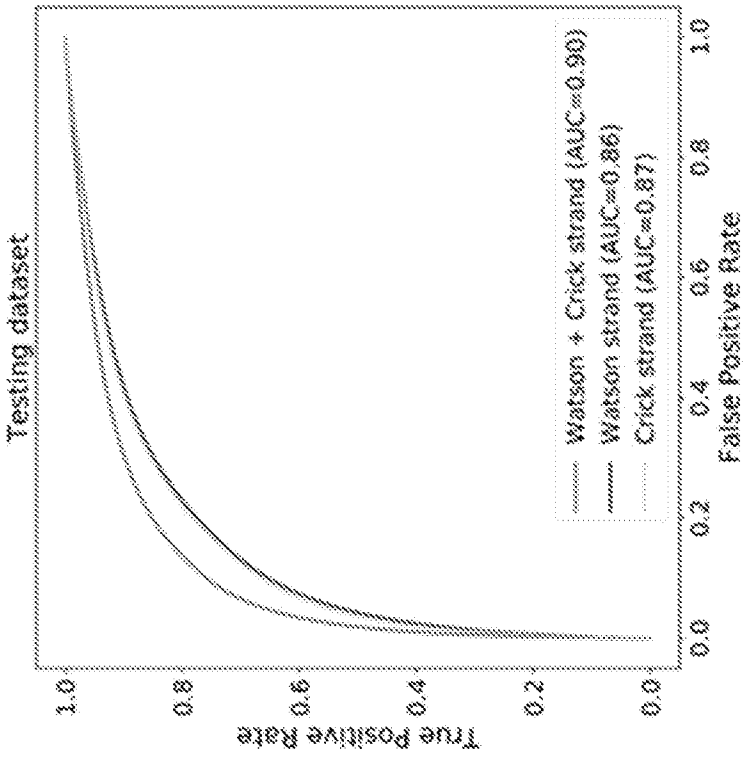


FIG. 20B

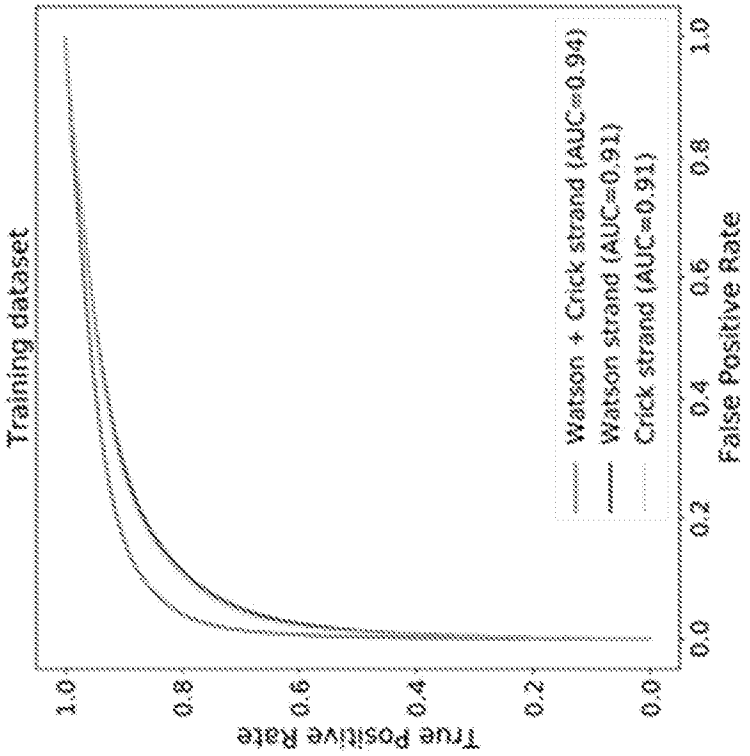


FIG. 20A

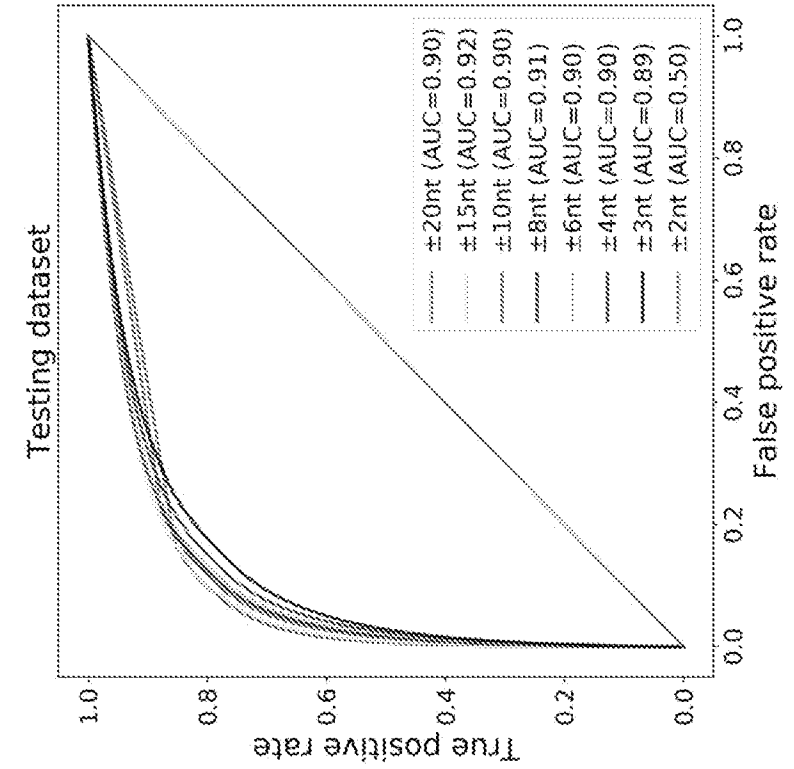


FIG. 21B

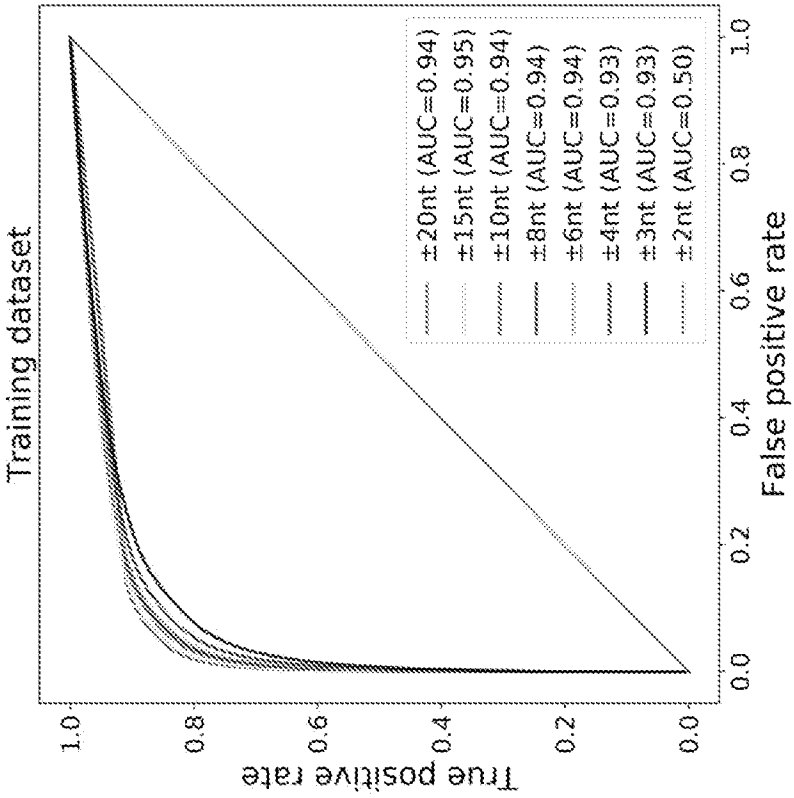


FIG. 21A

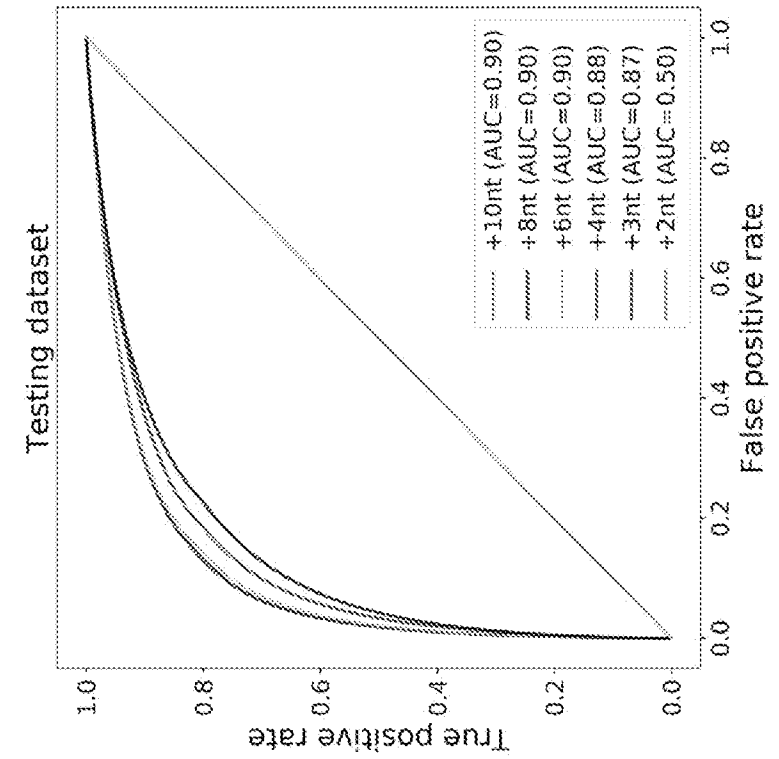


FIG. 22B

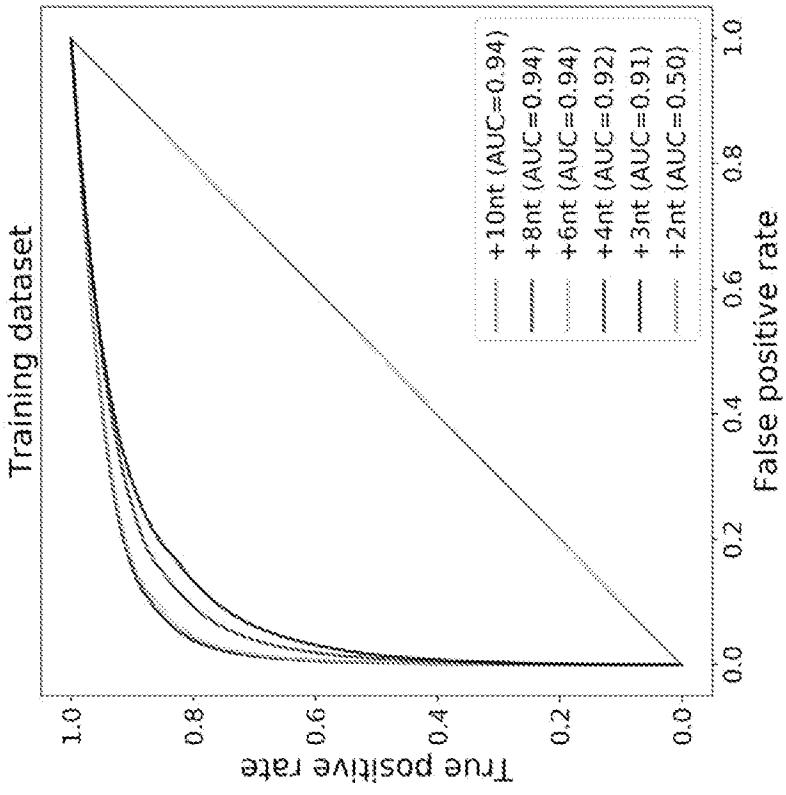


FIG. 22A

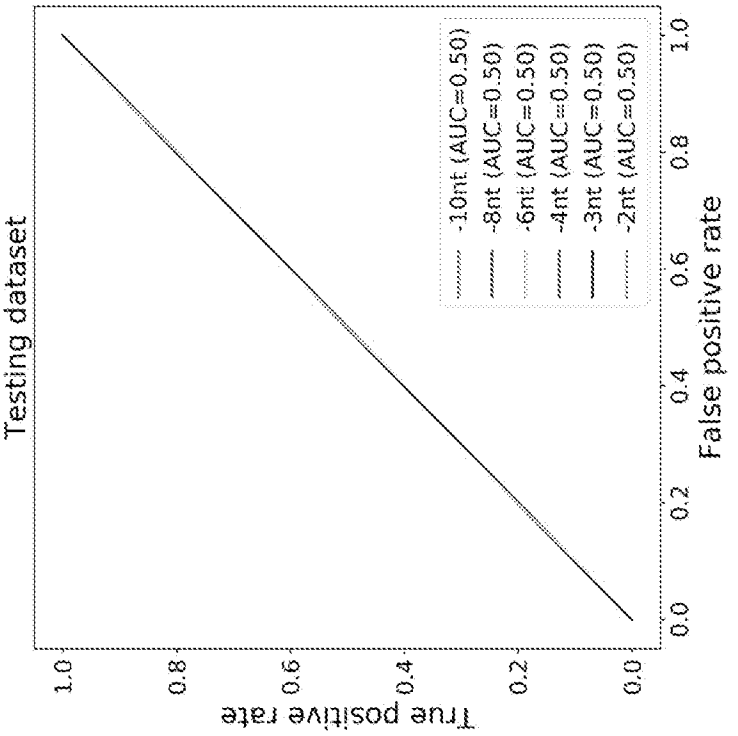


FIG. 23B

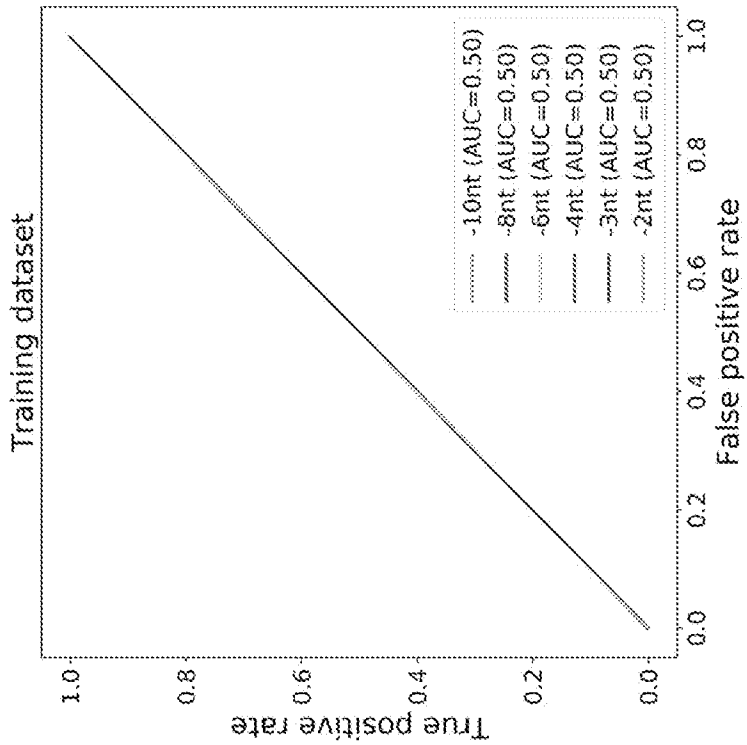


FIG. 23A

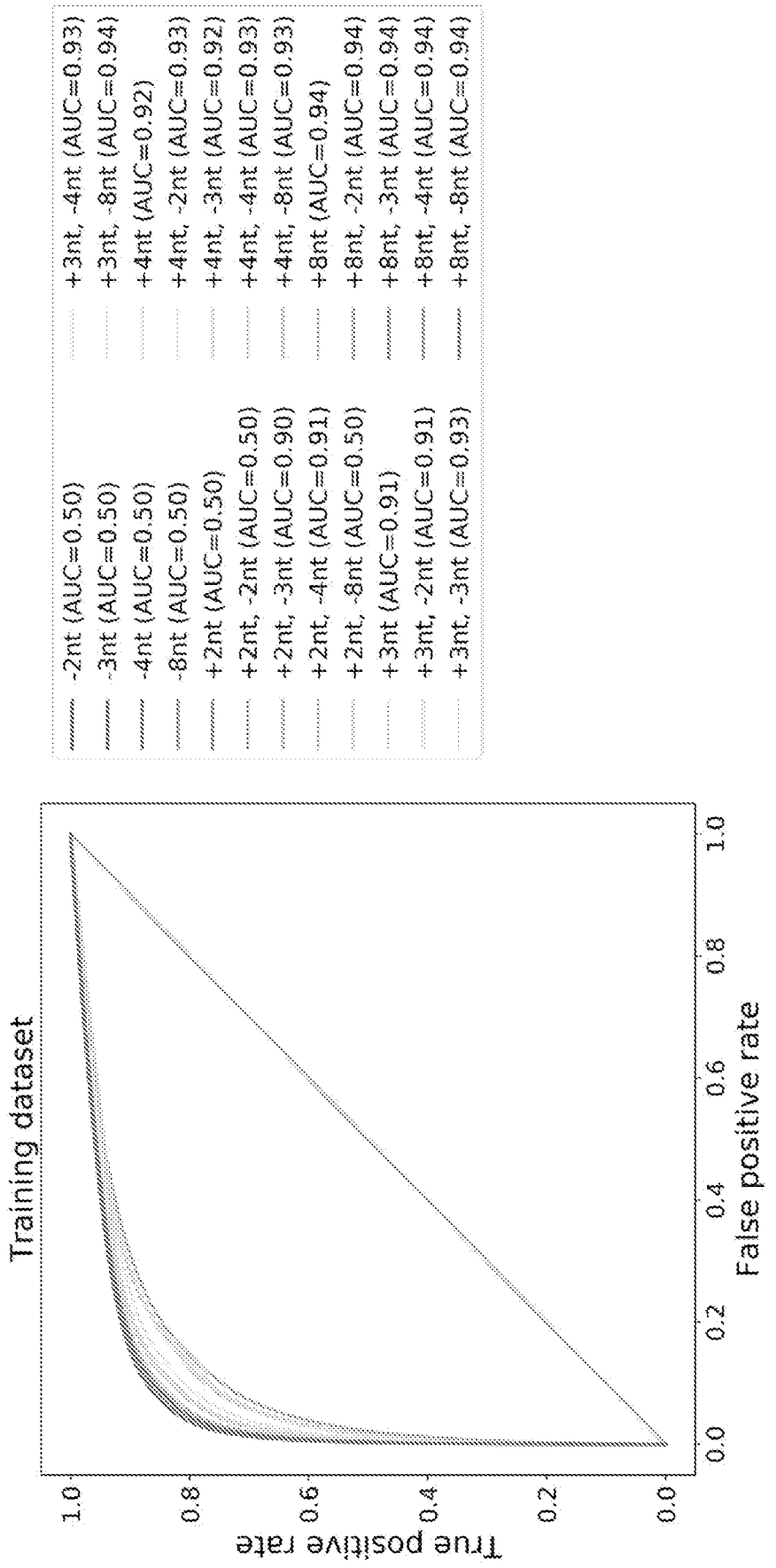

Training dataset
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
-2nt (AUC=0.50)
-3nt (AUC=0.50)
-4nt (AUC=0.50)
-8nt (AUC=0.50)
+2nt (AUC=0.50)
+2nt, -2nt (AUC=0.50)
+2nt, -3nt (AUC=0.90)
+2nt, -4nt (AUC=0.91)
+2nt, -8nt (AUC=0.50)
+3nt (AUC=0.91)
+3nt, -2nt (AUC=0.91)
+3nt, -3nt (AUC=0.93)
+3nt, -4nt (AUC=0.93)
+3nt, -8nt (AUC=0.94)
+4nt (AUC=0.92)
+4nt, -2nt (AUC=0.93)
+4nt, -3nt (AUC=0.92)
+4nt, -4nt (AUC=0.93)
+4nt, -8nt (AUC=0.93)
+8nt (AUC=0.94)
+8nt, -2nt (AUC=0.94)
+8nt, -3nt (AUC=0.94)
+8nt, -4nt (AUC=0.94)
+8nt, -8nt (AUC=0.94)

FIG. 24

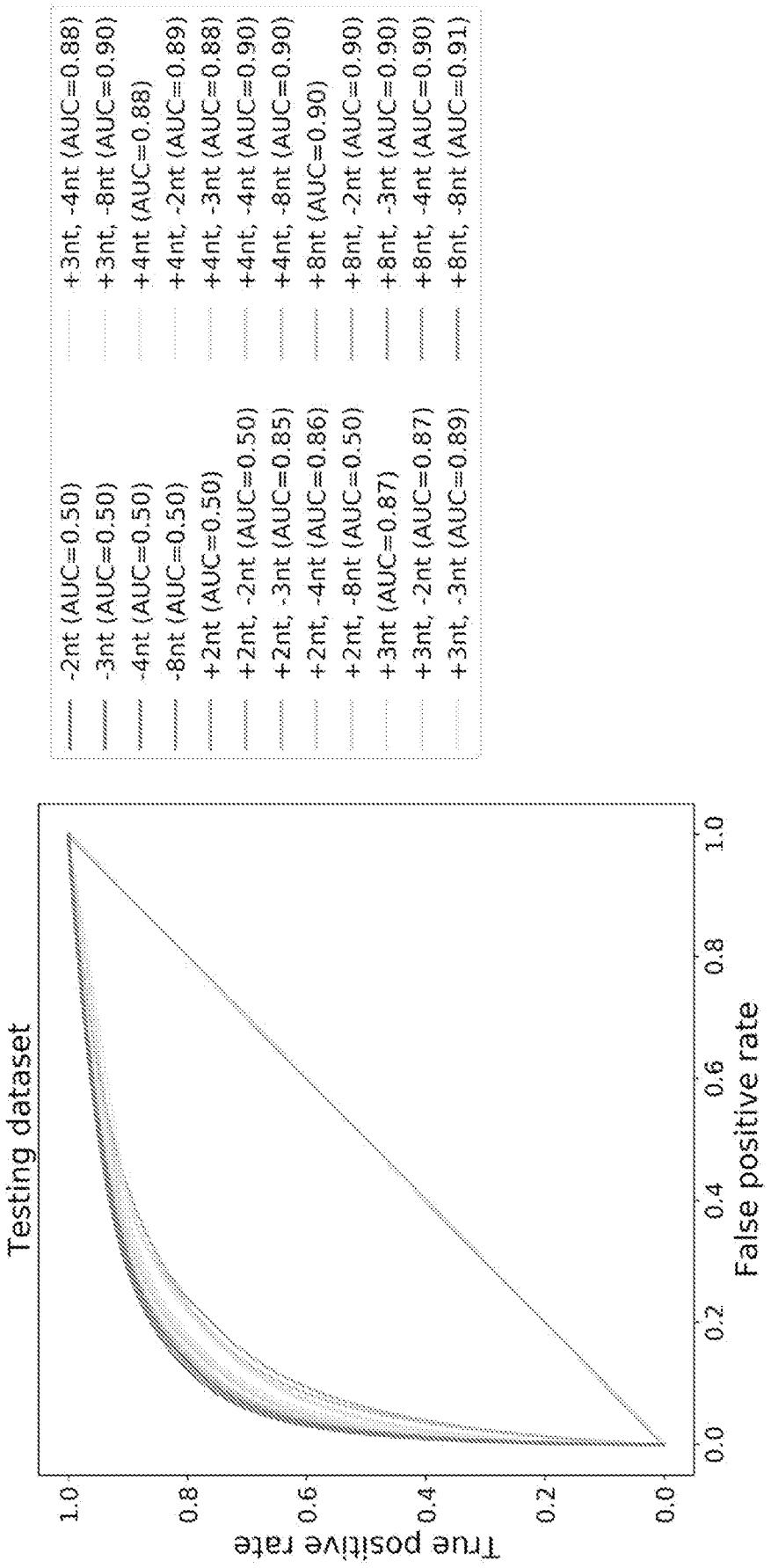

Testing dataset
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
-2nt (AUC=0.50)
-3nt (AUC=0.50)
-4nt (AUC=0.50)
-8nt (AUC=0.50)
+2nt (AUC=0.50)
+2nt, -2nt (AUC=0.50)
+2nt, -3nt (AUC=0.85)
+2nt, -4nt (AUC=0.86)
+2nt, -8nt (AUC=0.50)
+3nt (AUC=0.87)
+3nt, -2nt (AUC=0.87)
+3nt, -3nt (AUC=0.89)
+3nt, -4nt (AUC=0.88)
+3nt, -8nt (AUC=0.90)
+4nt (AUC=0.88)
+4nt, -2nt (AUC=0.89)
+4nt, -3nt (AUC=0.88)
+4nt, -4nt (AUC=0.90)
+4nt, -8nt (AUC=0.90)
+8nt (AUC=0.90)
+8nt, -2nt (AUC=0.90)
+8nt, -3nt (AUC=0.90)
+8nt, -4nt (AUC=0.90)
+8nt, -8nt (AUC=0.91)

FIG. 25

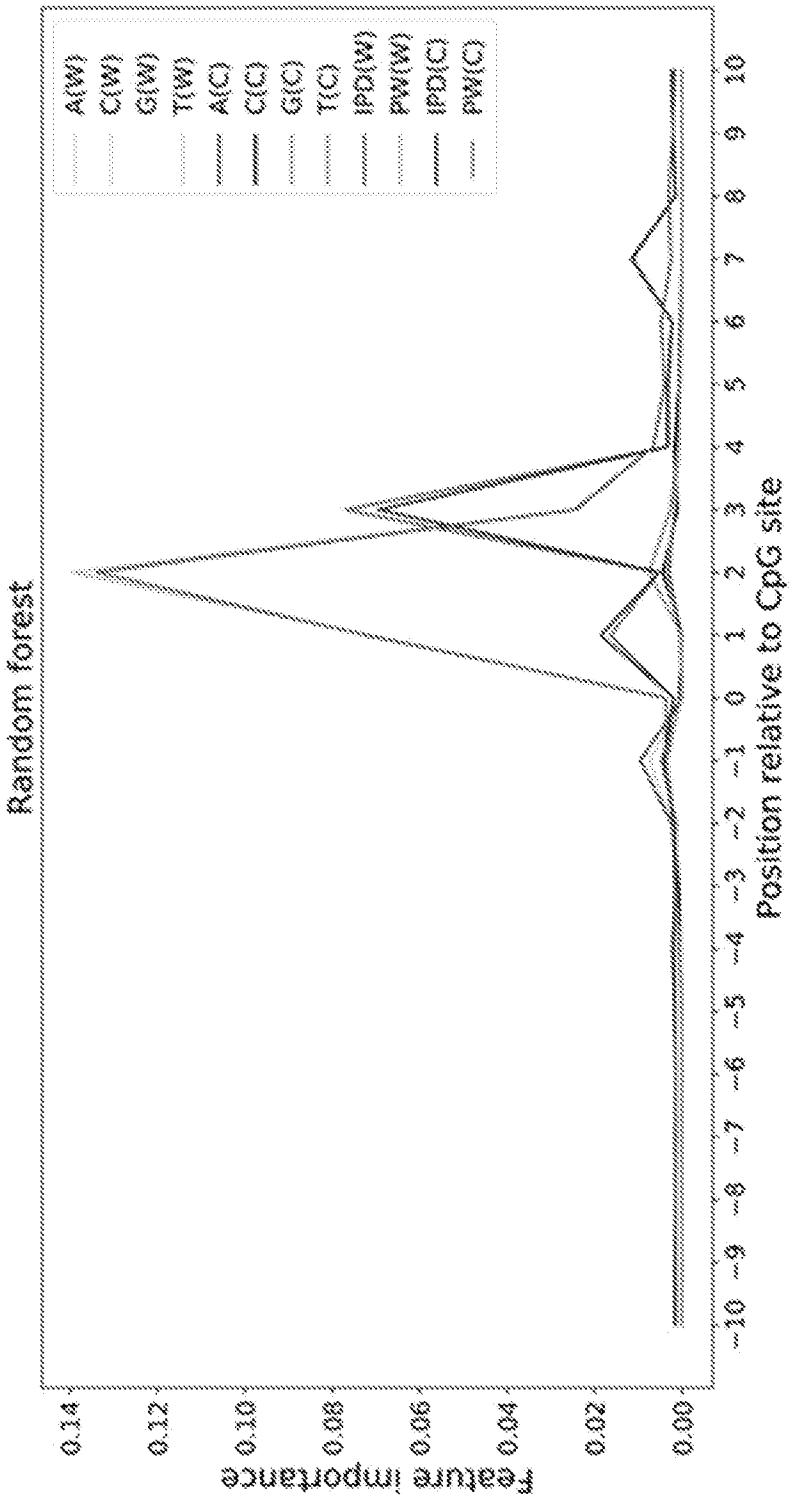

Random forest
A(W)
C(W)
G(W)
T(W)
A(C)
C(C)
G(C)
T(C)
IPD(W)
PW(W)
IPD(C)
PW(C)
feature importance
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10
Position relative to CpG site

FIG. 26

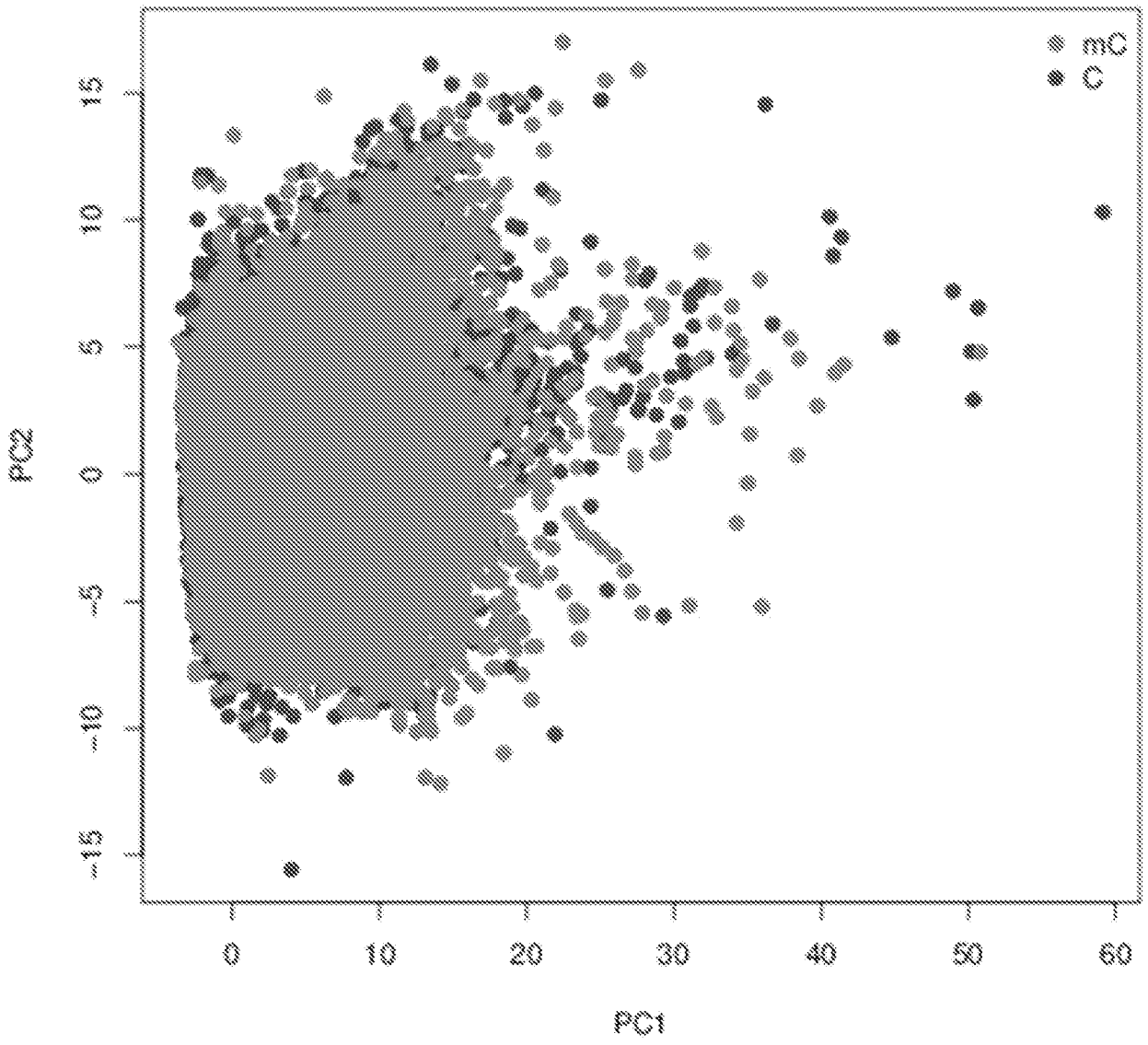


FIG. 28

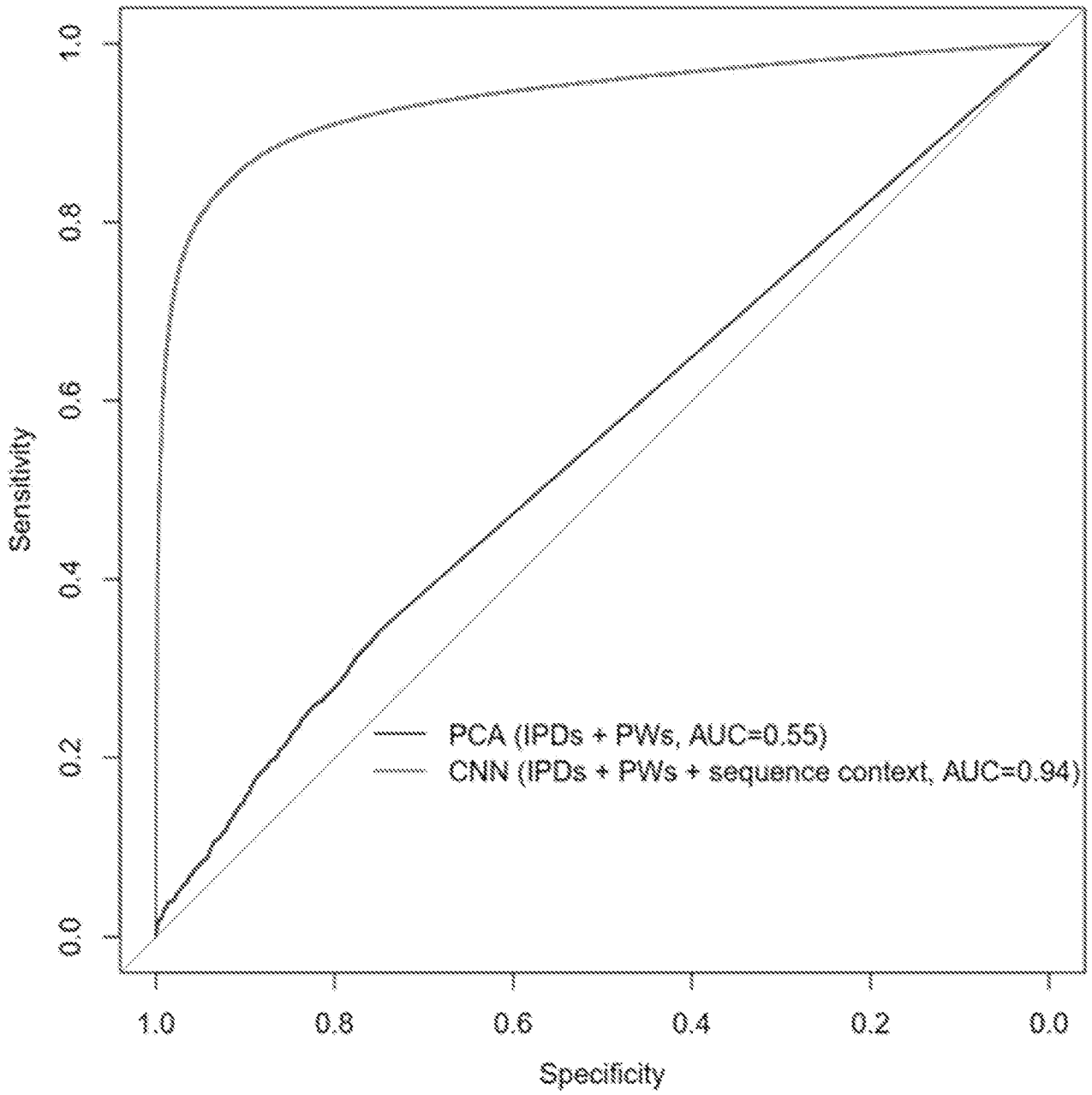

Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
PCA (IPDs + PWs, AUC=0.55)
CNN (IPDs + PWs + sequence context, AUC=0.94)
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

FIG. 29

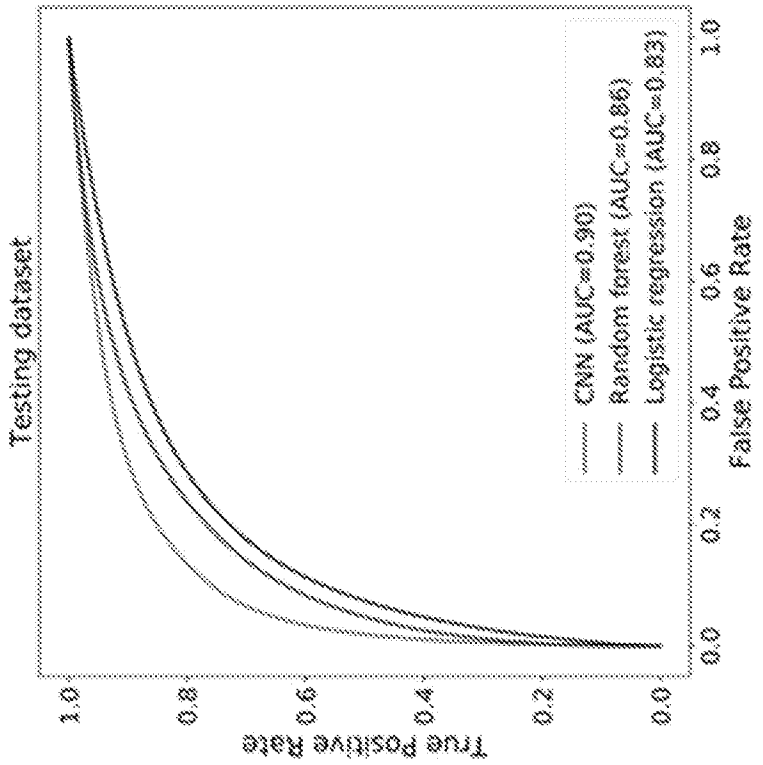


FIG. 30B

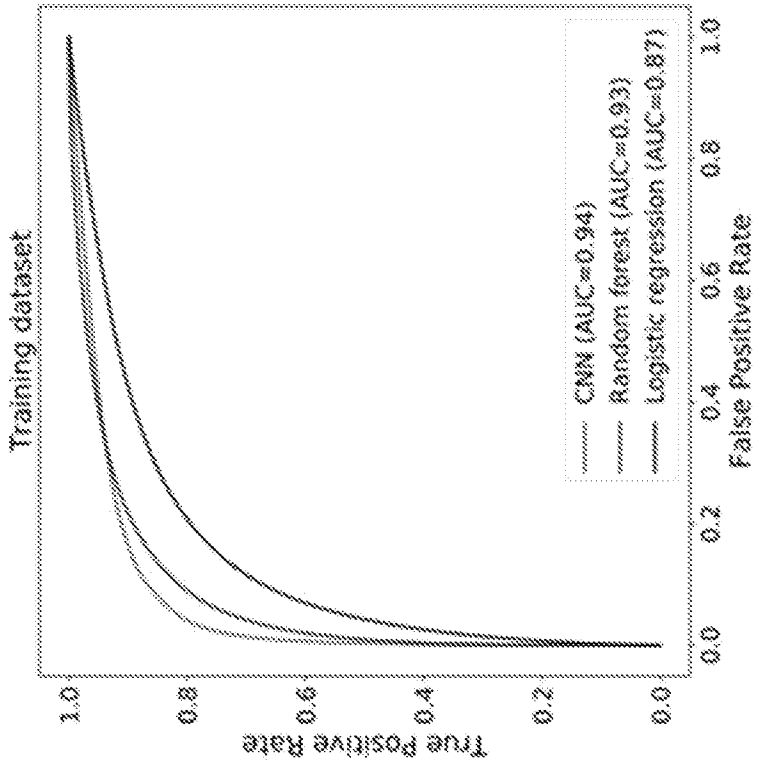
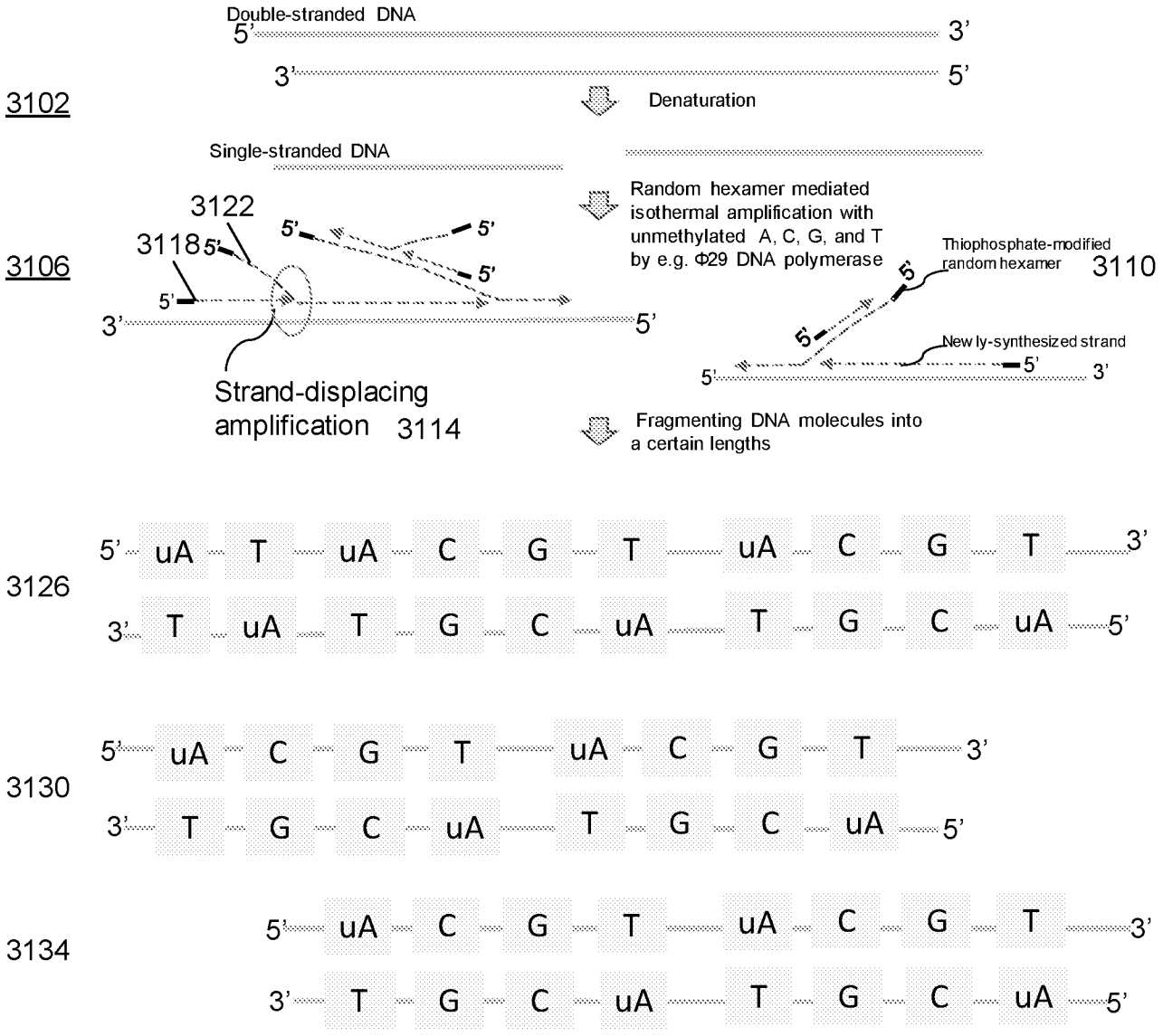
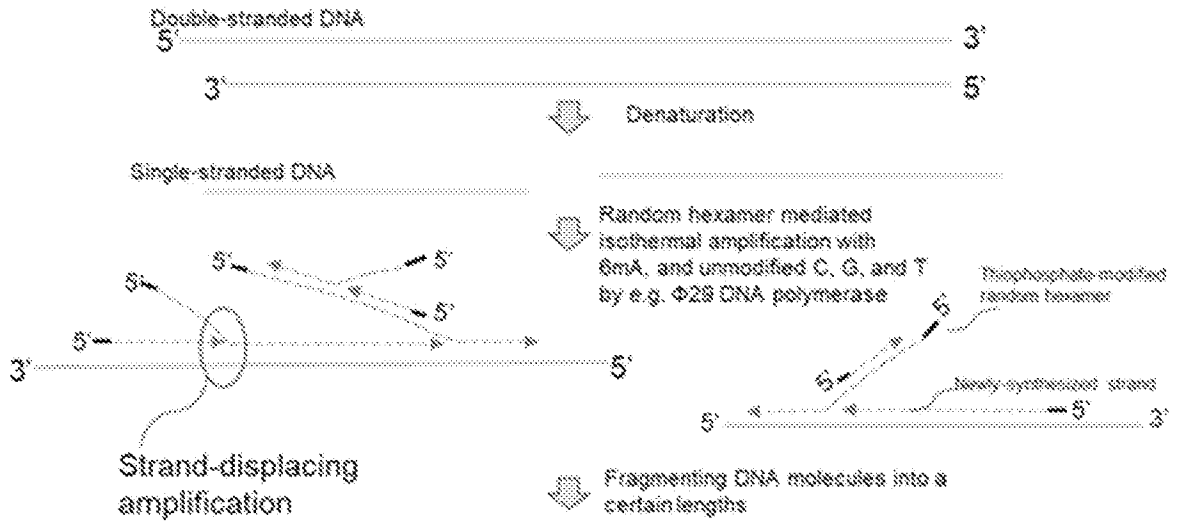


FIG. 30A



Whole-genome amplified DNA products

FIG. 31A



(I)

5' mA T mA C G T mA C G T 3'

3' T mA T G C mA T G C mA 5'

(II)

5' uA C G T mA C G T 3'

3' T G C mA T G C mA 5'

(III)

5' uA C G T uA C G T 3'

3' T G C mA T G C mA 5'

(IV)

5' uA C G T uA C G T 3'

3' T G C uA T G C uA 5'

Whole-genome amplified DNA products

FIG. 31B

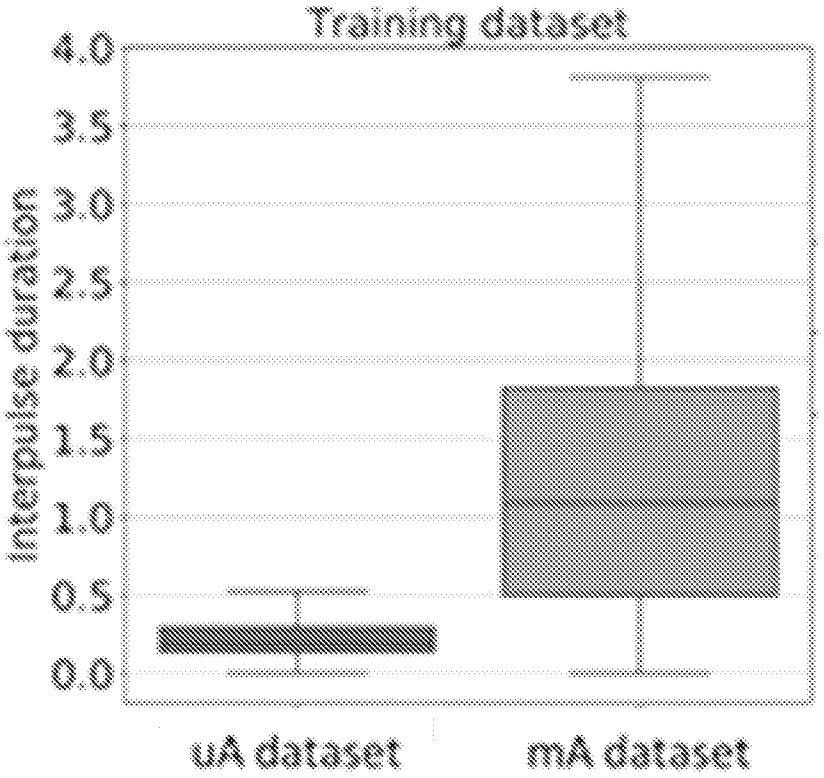


FIG. 32A

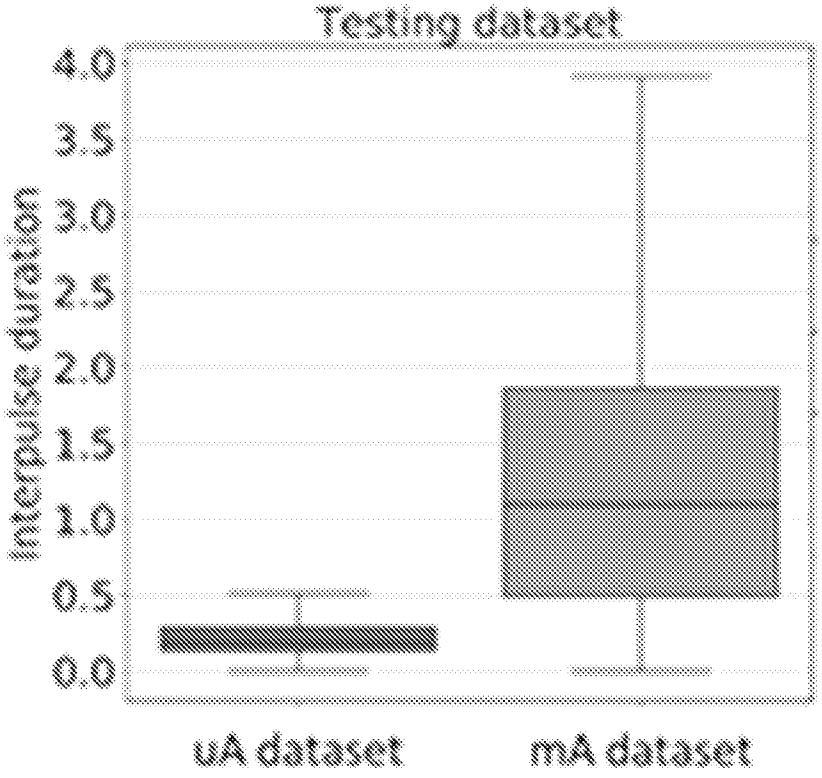


FIG. 32B

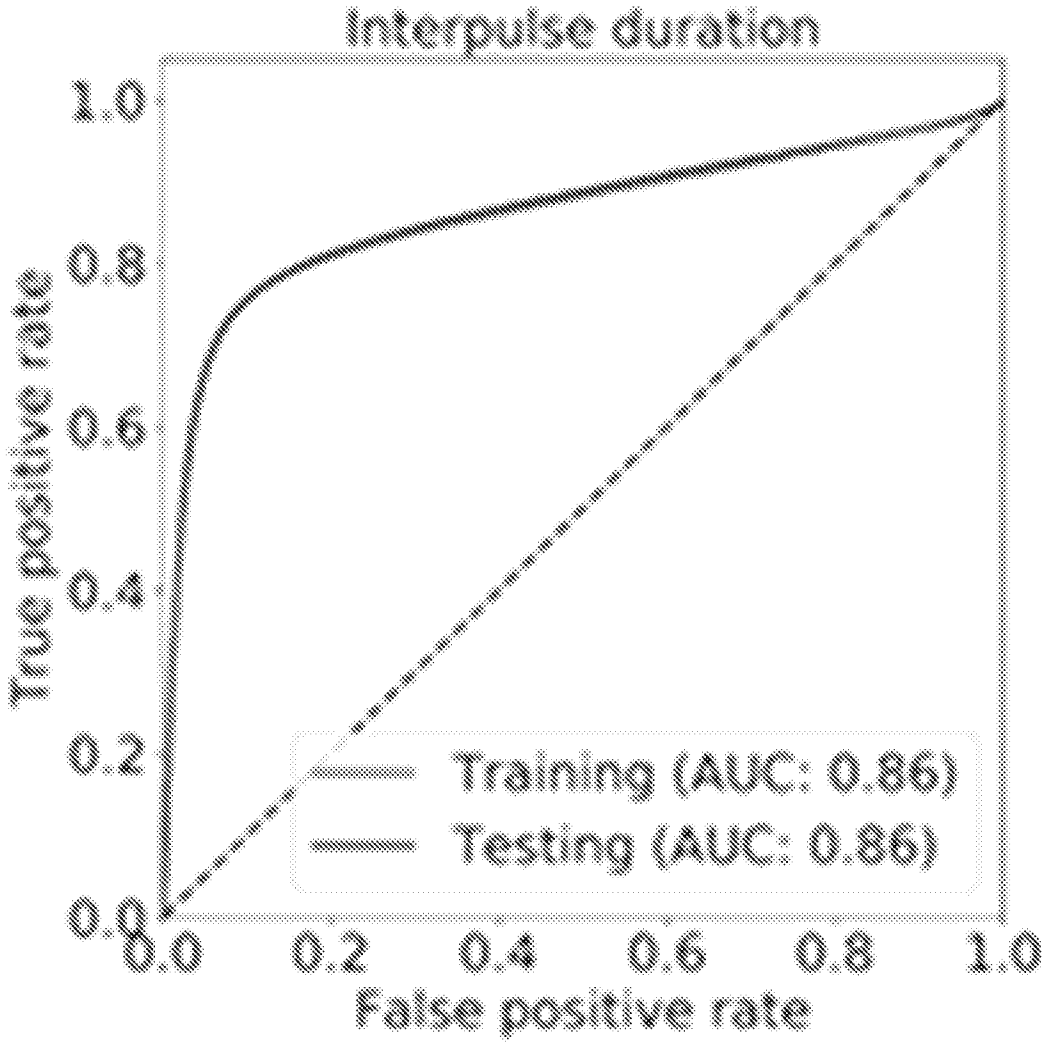

Interpulse duration
True positive rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
Training (AUC: 0.86)
Testing (AUC: 0.86)

FIG. 32C

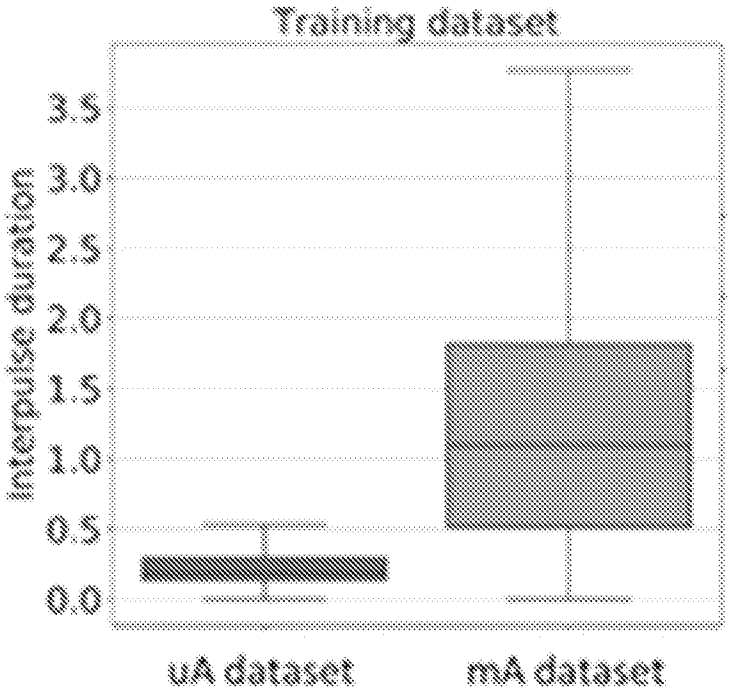


FIG. 33A

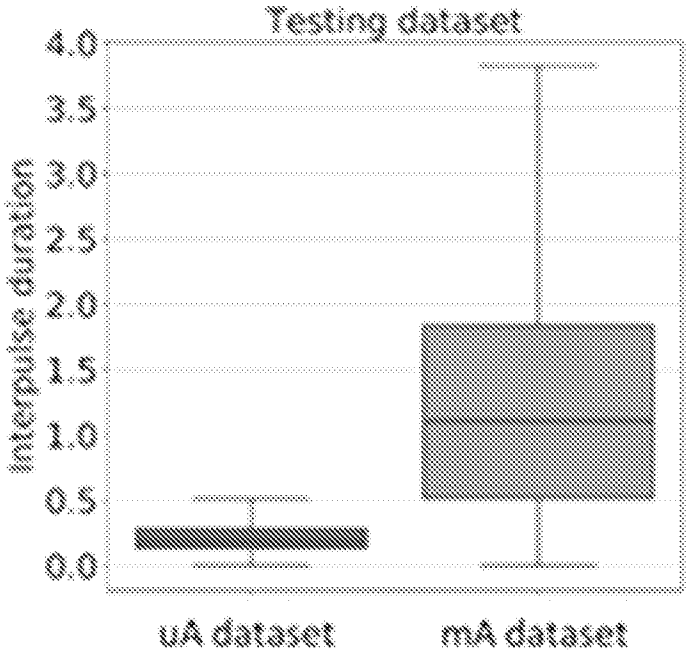


FIG. 33B

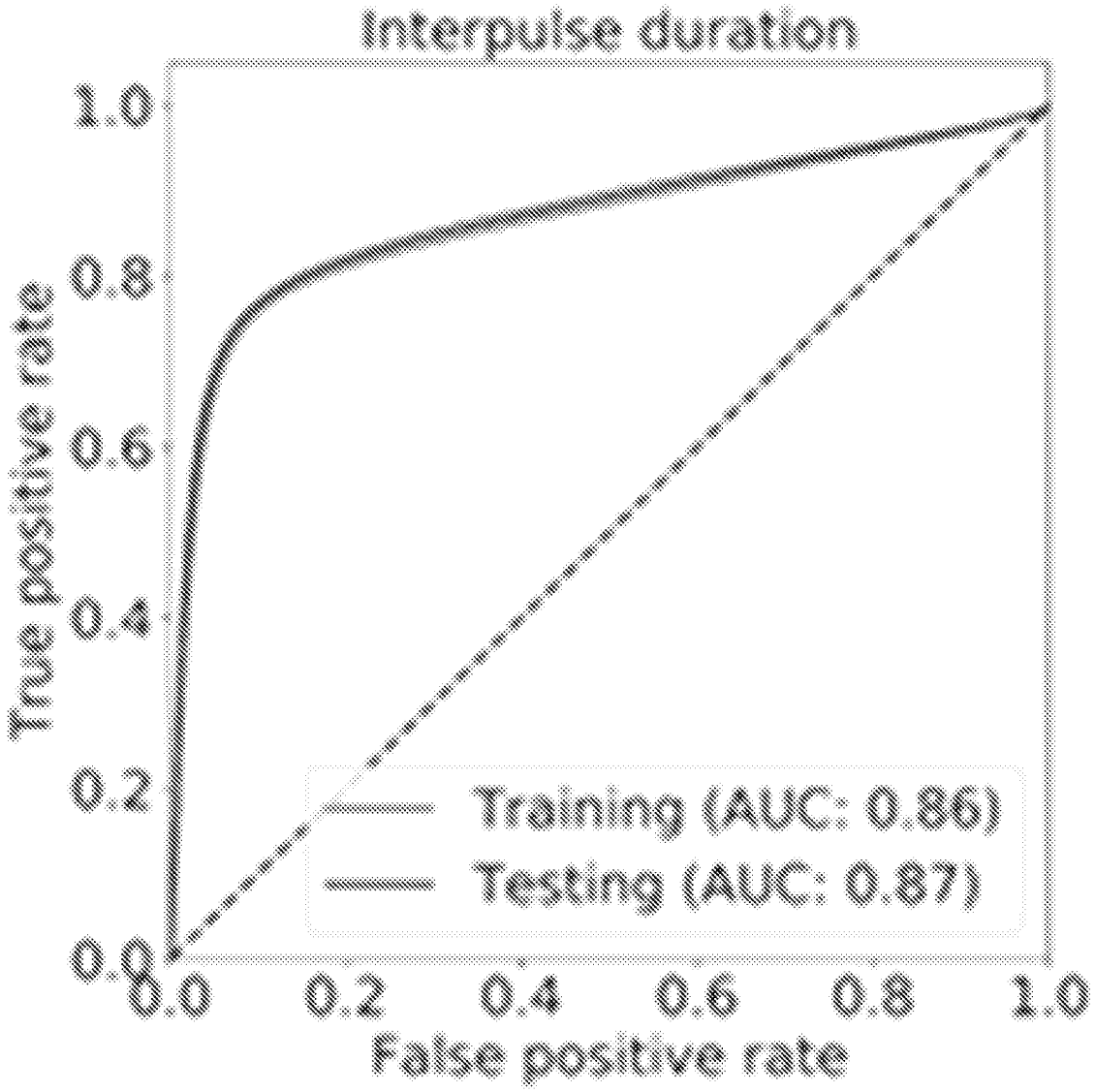

Interpulse duration
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Training (AUC: 0.86)
Testing (AUC: 0.87)

FIG. 33C

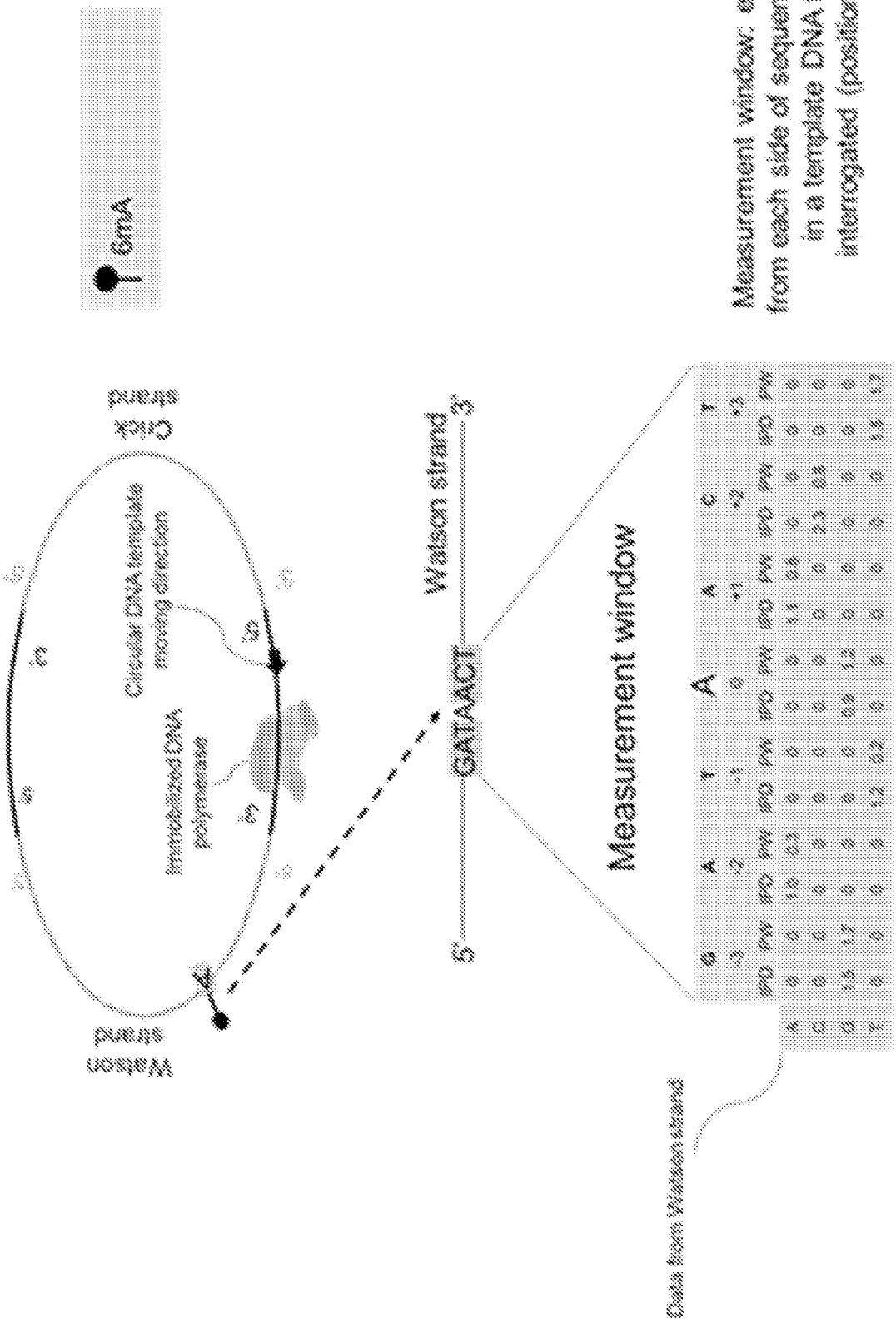

6mA
Measurement window, e.g. 3 bases from each side of sequenced A base in a template DNA that is interrogated (position of '0')
Watson strand
Crick strand
Circular DNA template moving direction
Immobilized DNA polymerase
Watson strand 3'
GATAACT
5'
Measurement window
Data from Watson strand

FIG. 34

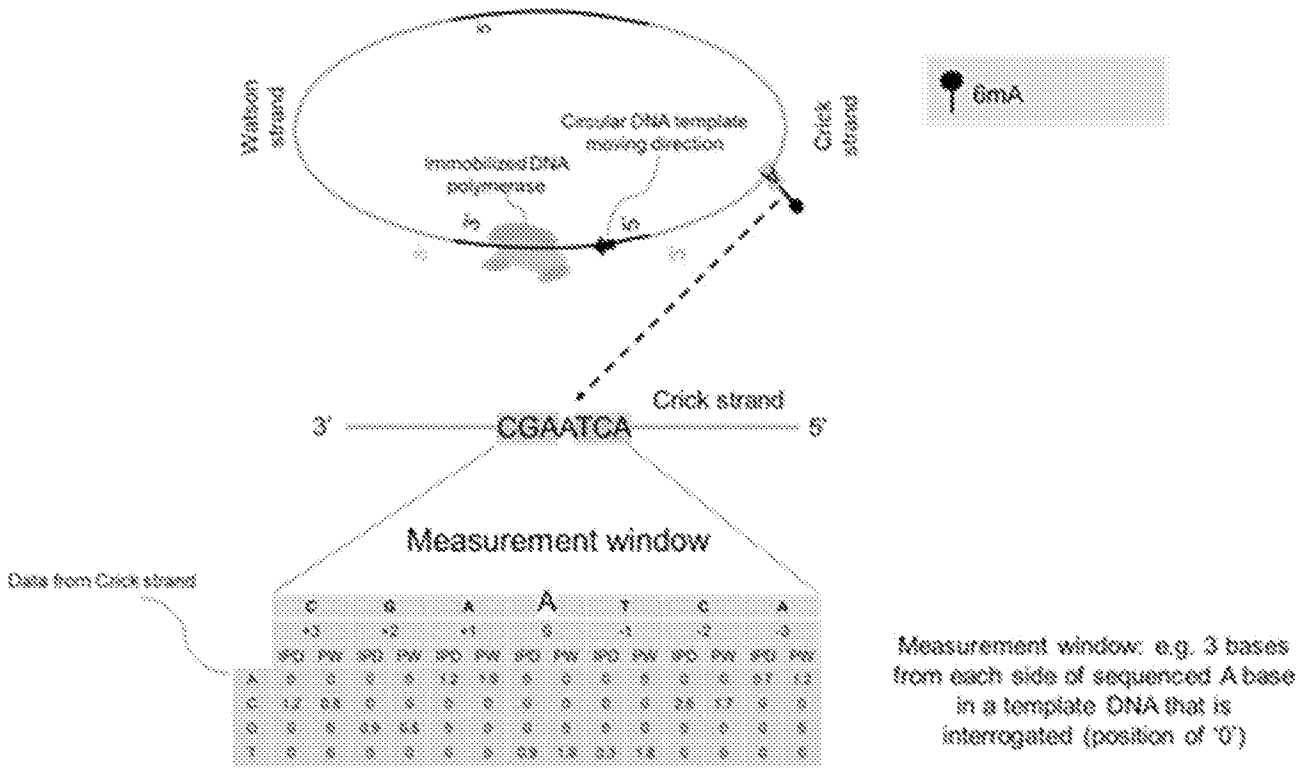

Watson strand
Crick strand
Circular DNA template moving direction
Immobilized DNA polymerase
6mA
3'
CGAATCA
Crick strand
5'
Measurement window
Data from Crick strand
Measurement window: e.g. 3 bases from each side of sequenced A base in a template DNA that is interrogated (position of '0')

FIG.35

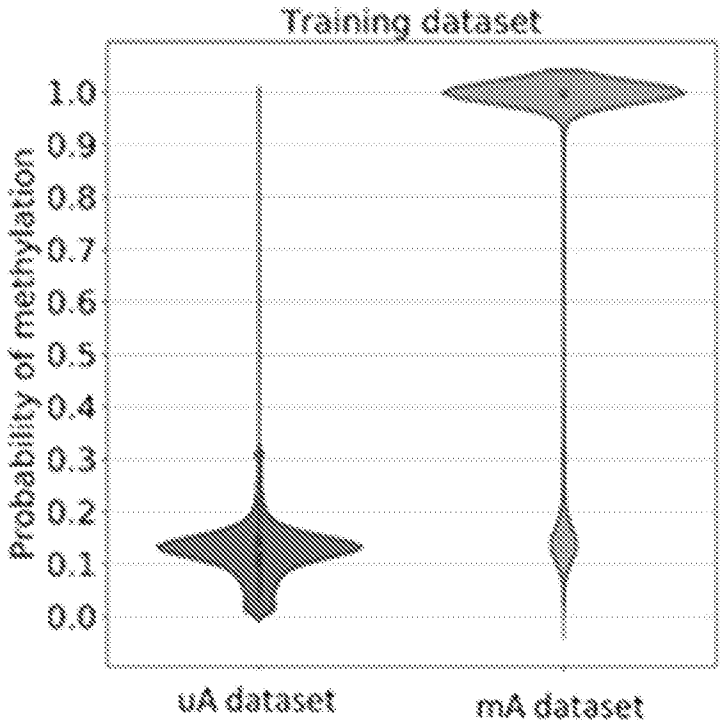


FIG. 36A

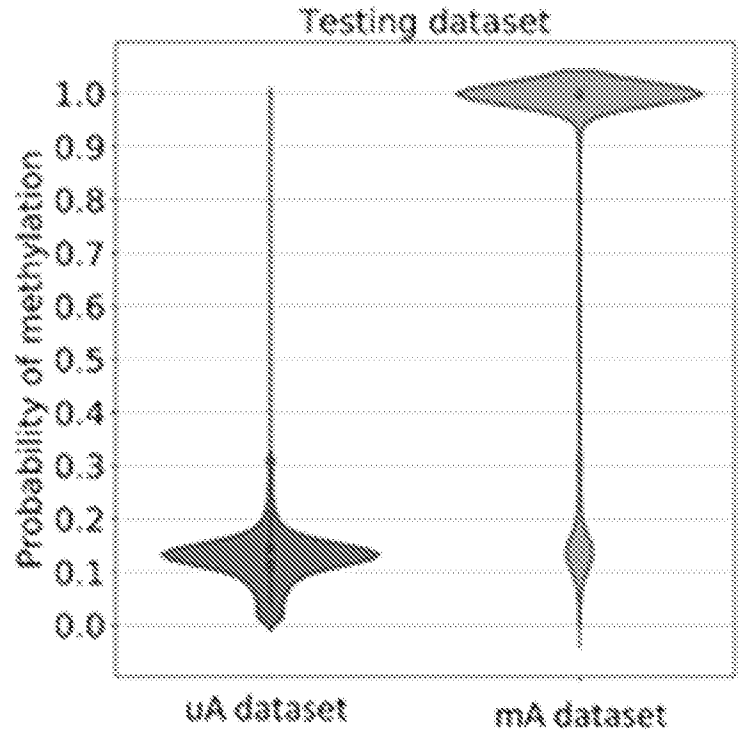


FIG. 36B

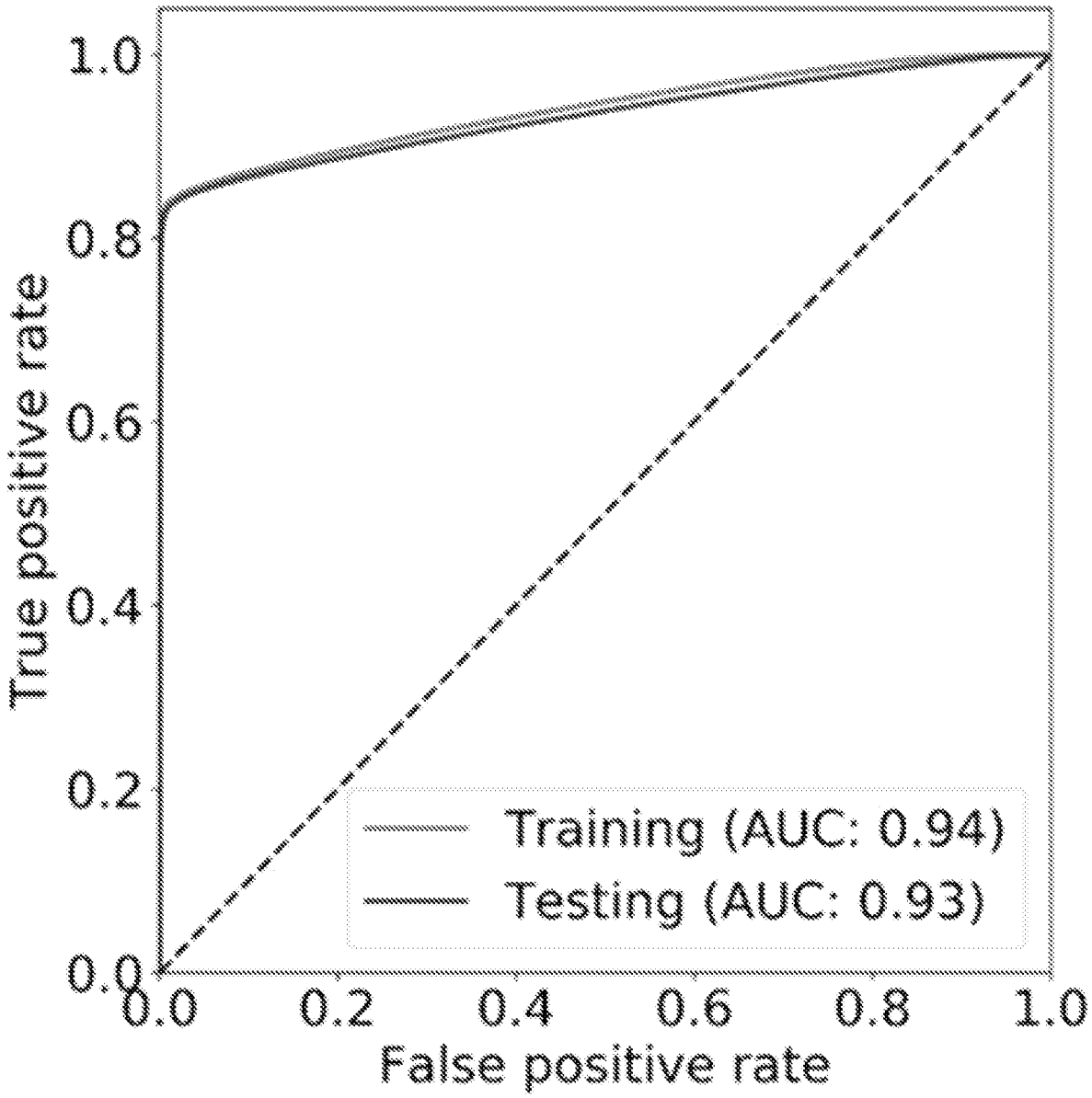

True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Training (AUC: 0.94)
Testing (AUC: 0.93)

FIG. 37

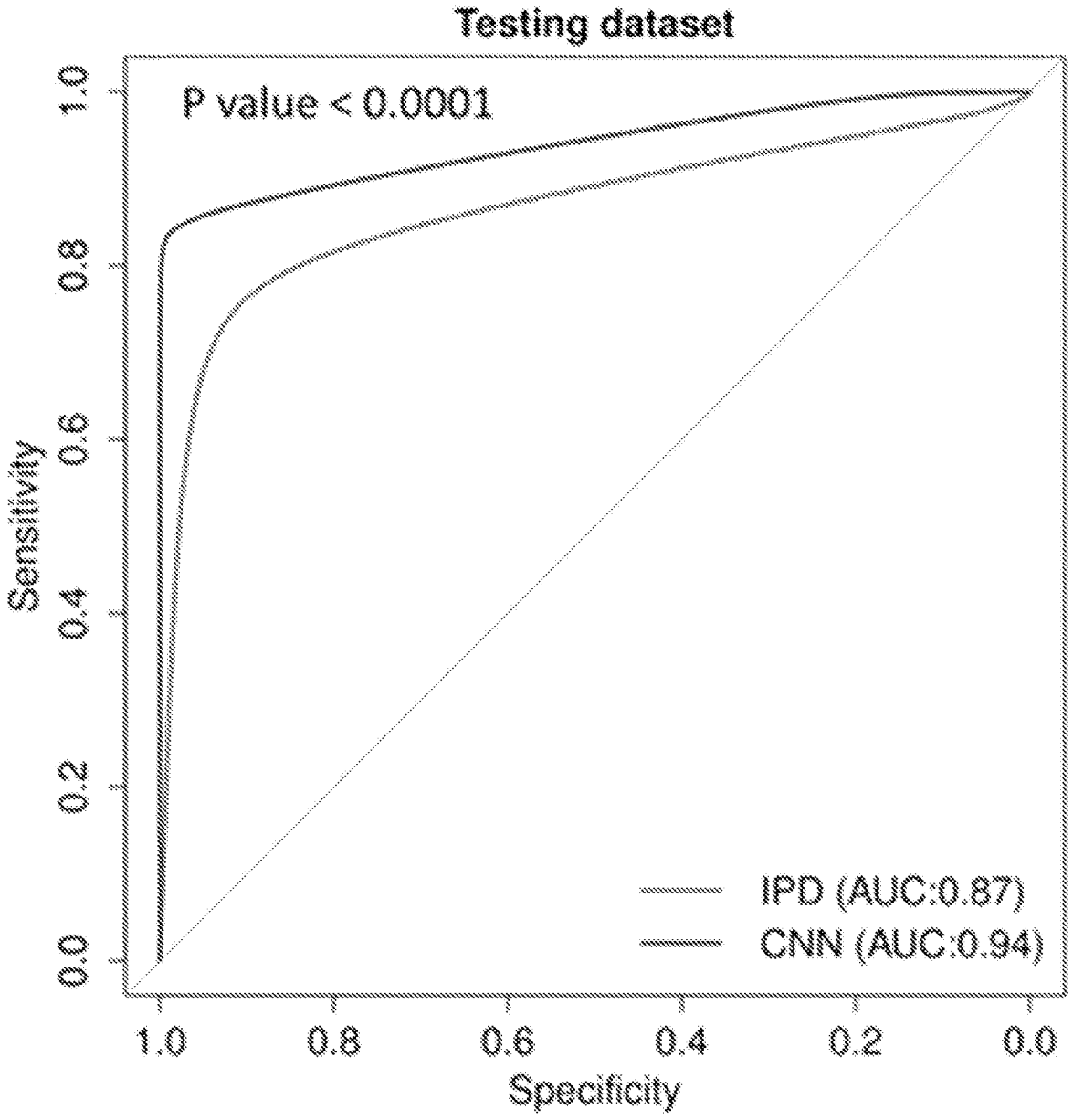

Testing dataset
P value < 0.0001
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
IPD (AUC:0.87)
CNN (AUC:0.94)
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

FIG. 38

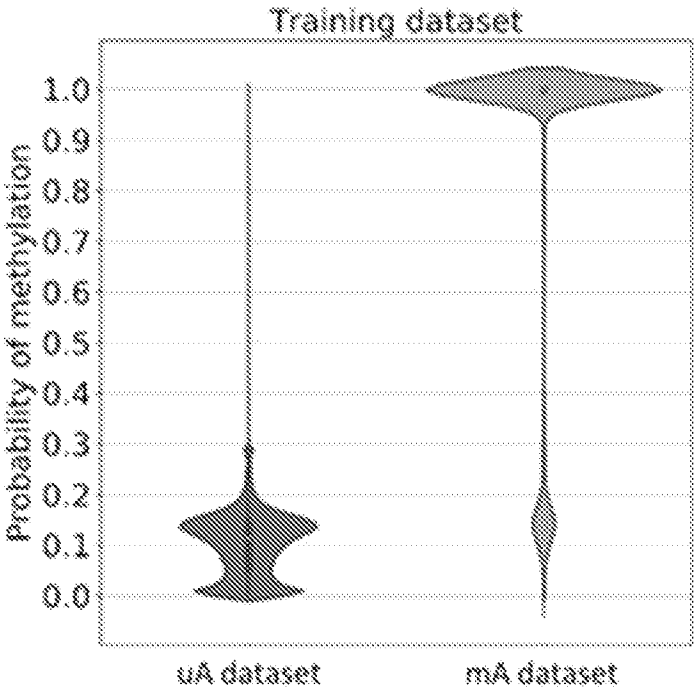


FIG. 39A

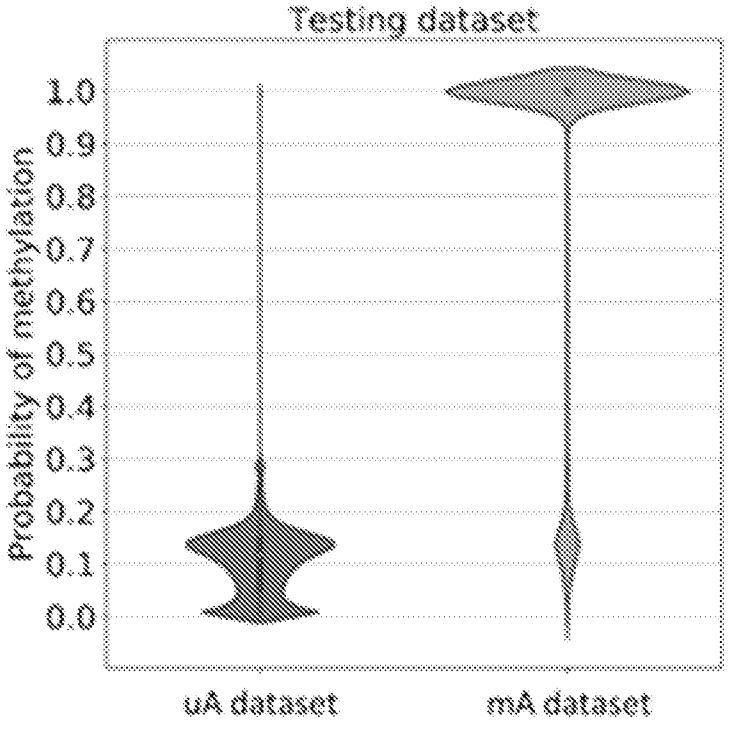


FIG. 39B

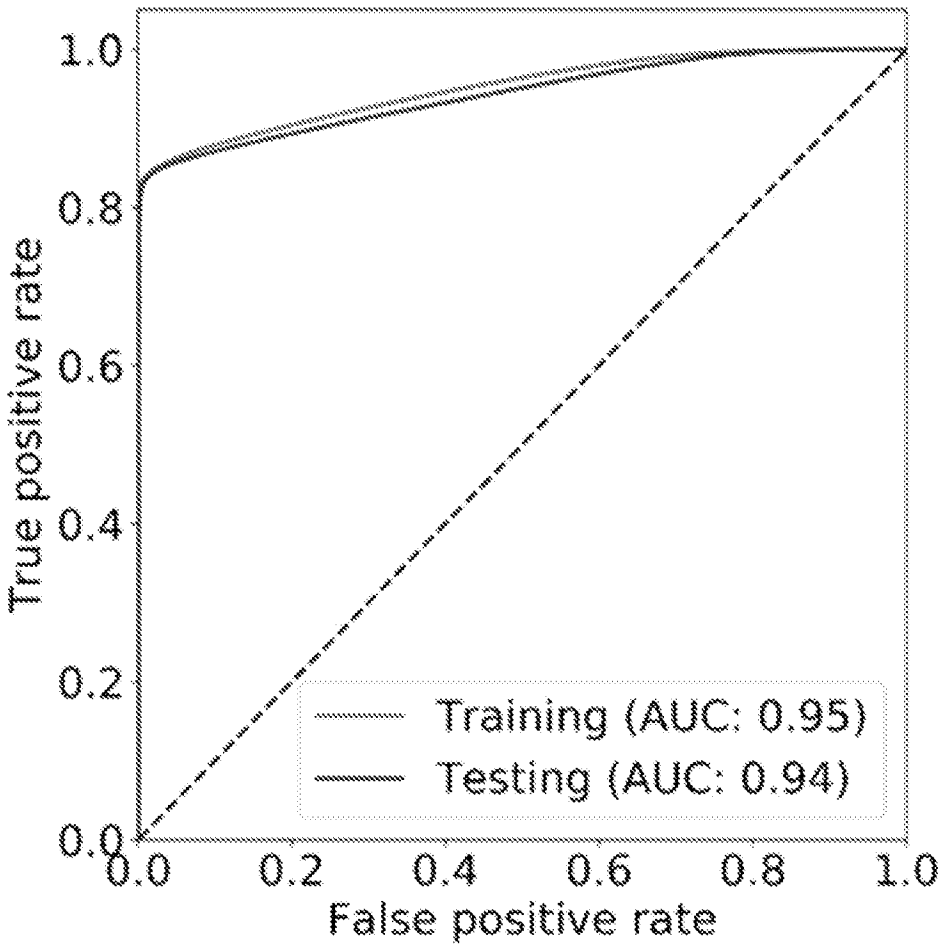

1.0
0.8
0.6
0.4
0.2
0.0
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
Training (AUC: 0.95)
Testing (AUC: 0.94)

FIG. 40

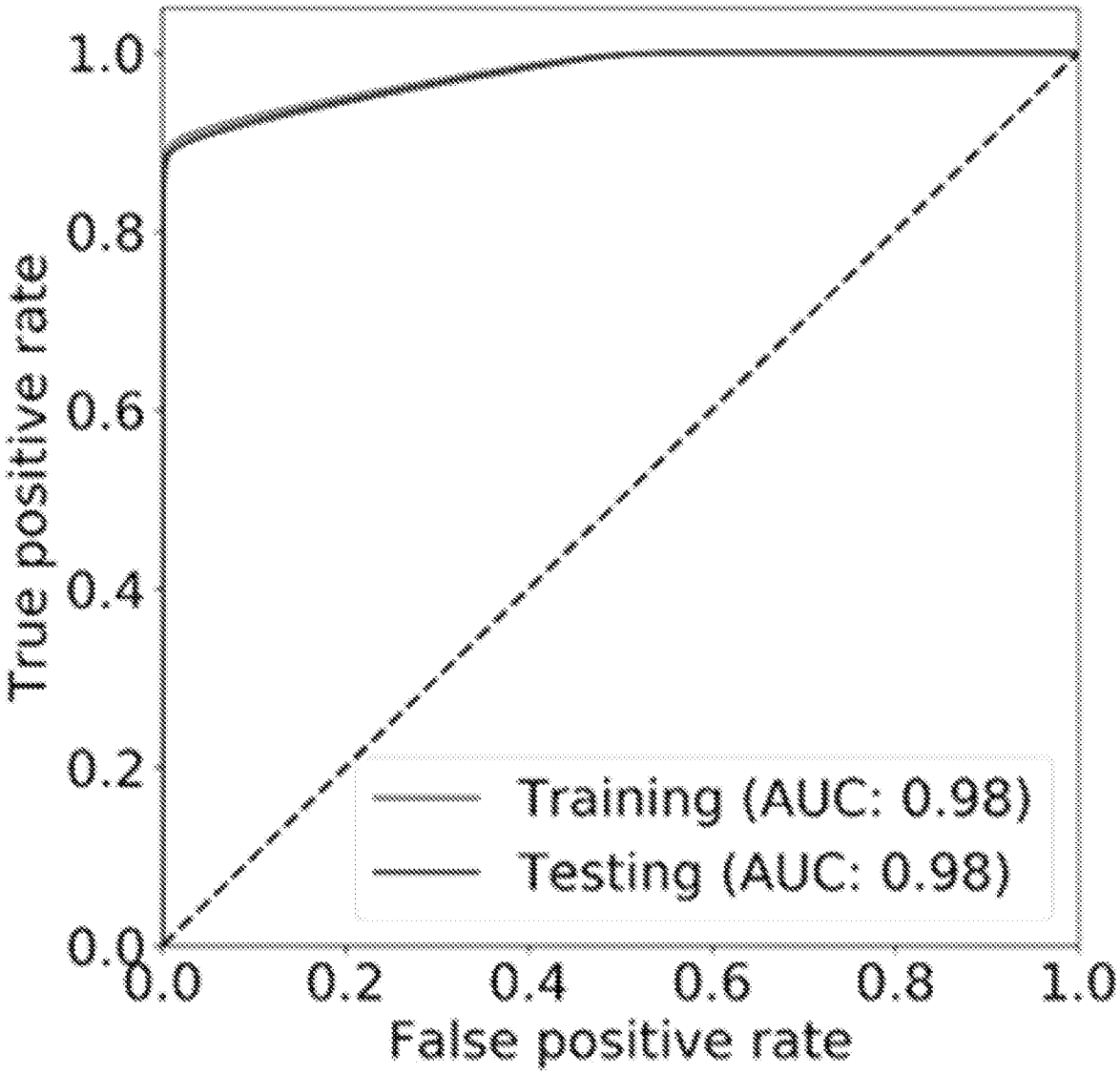

True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Training (AUC: 0.98)
Testing (AUC: 0.98)

FIG. 42

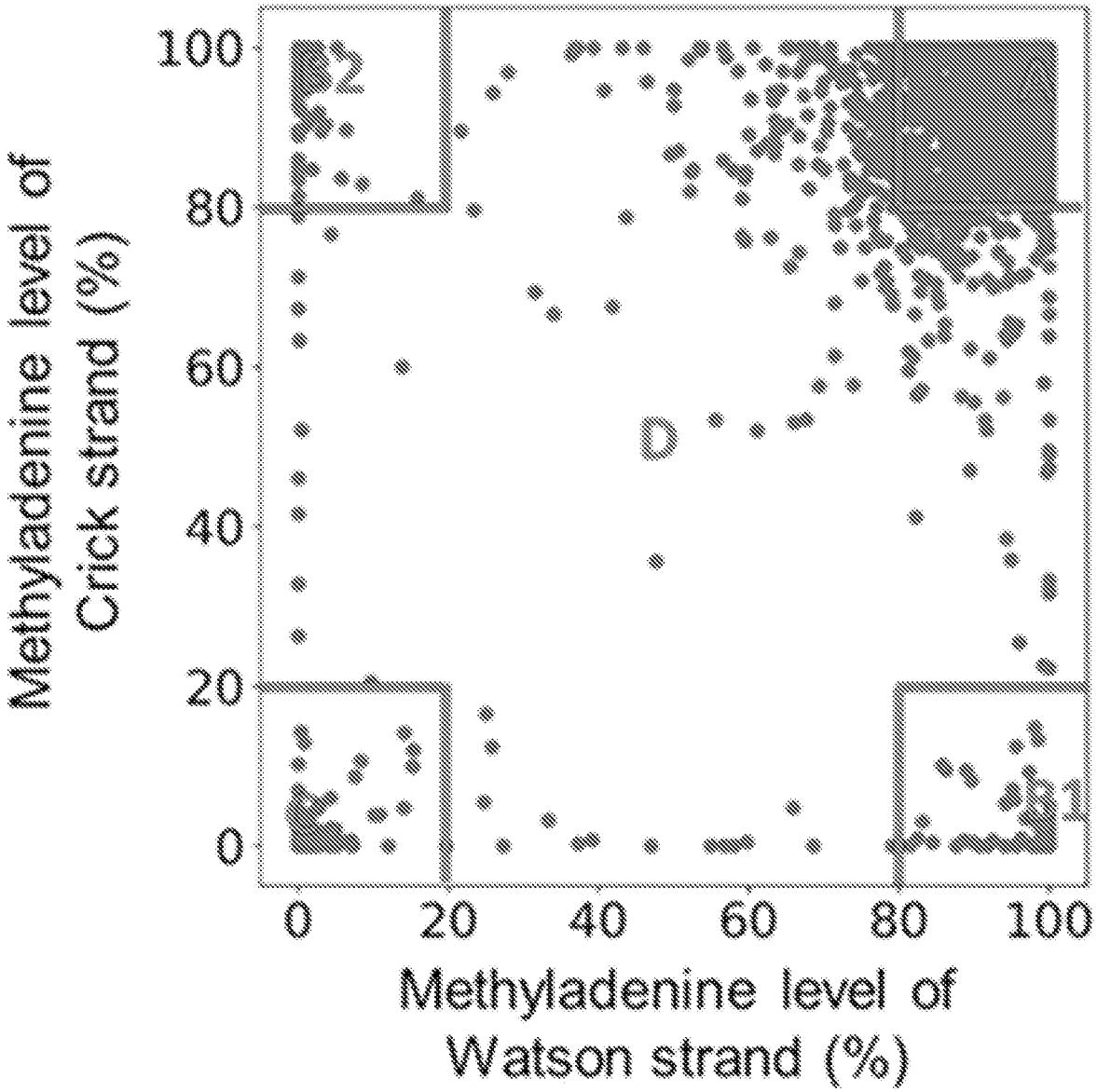

100
80
60
40
20
0
0
20
40
60
80
100
Methyladenine level of Crick strand (%)
Methyladenine level of Watson strand (%)
2
1

FIG.44

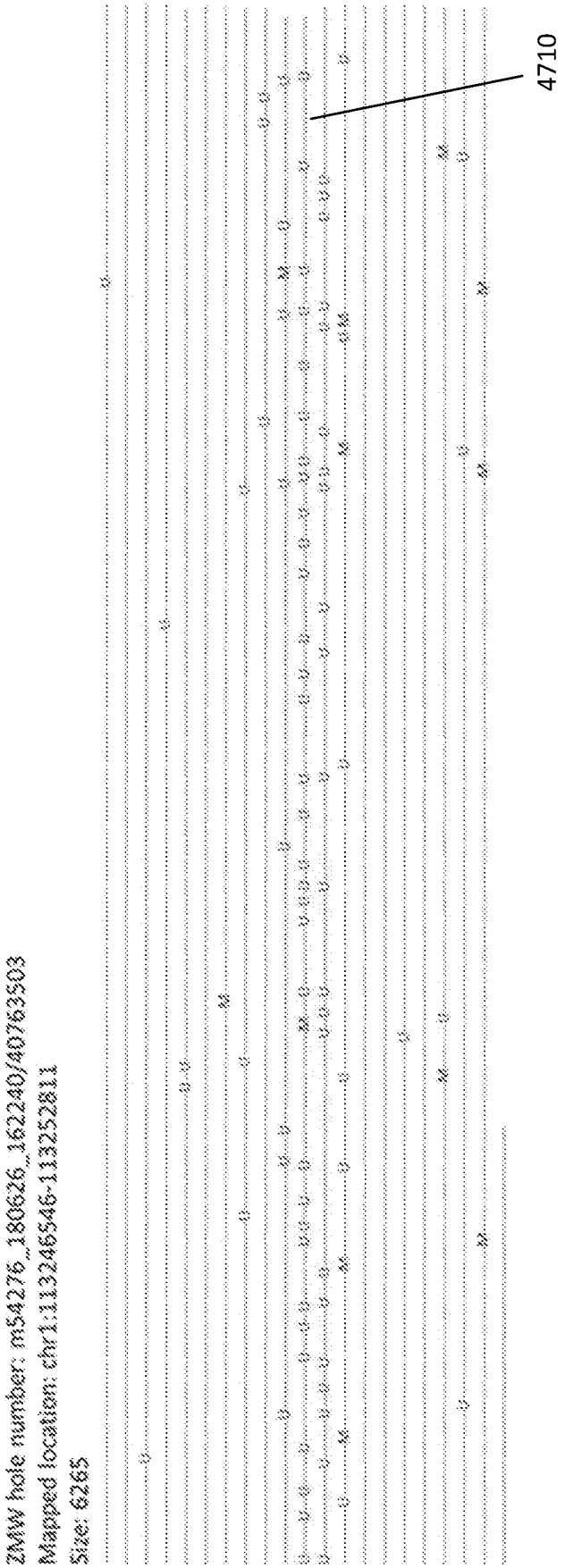

4710
ZMW hole number: m54276_180626_162240/40763503
Mapped location: chr1:113246546-113252611
Size: 6265

FIG. 47

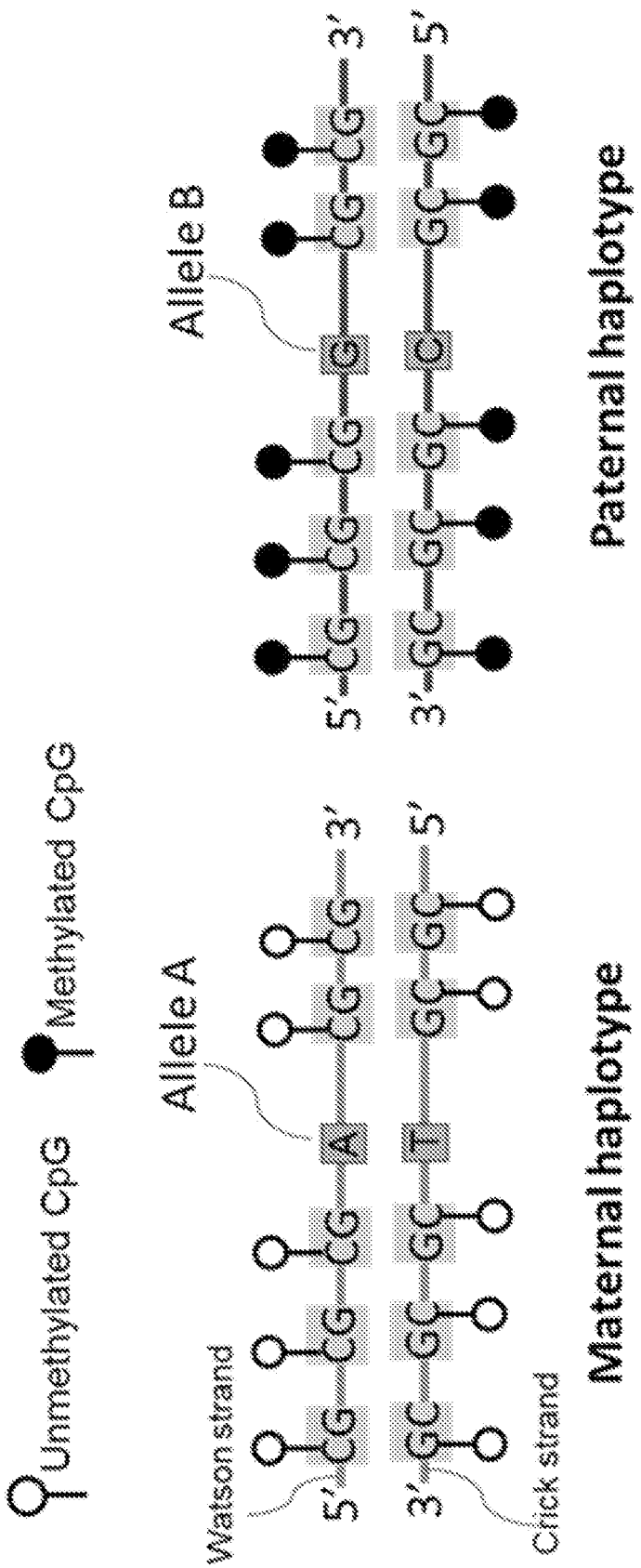

Methylation patterns present in a paternally imprinted region
Unmethylated CpG
Methylated CpG
Allele A
Allele B
Watson strand
Crick strand
5'
3'
A
T
Maternal haplotype
Paternal haplotype

FIG. 49

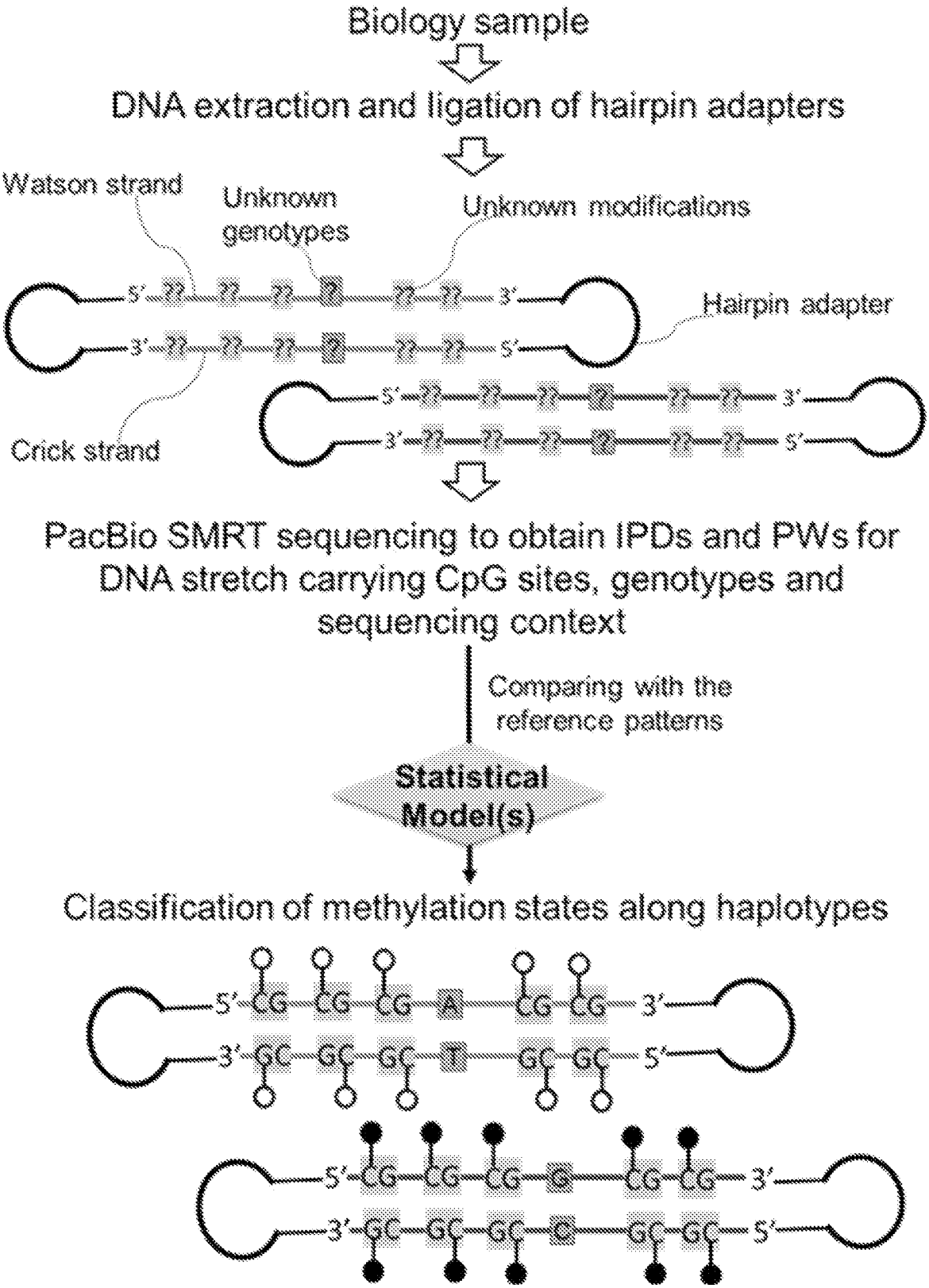

Biology sample
DNA extraction and ligation of hairpin adapters
Watson strand
Unknown genotypes
Unknown modifications
Hairpin adapter
Crick strand
PacBio SMRT sequencing to obtain IPDs and PWs for DNA stretch carrying CpG sites, genotypes and sequencing context
Comparing with the reference patterns
Statistical Model(s)
Classification of methylation states along haplotypes

FIG. 50

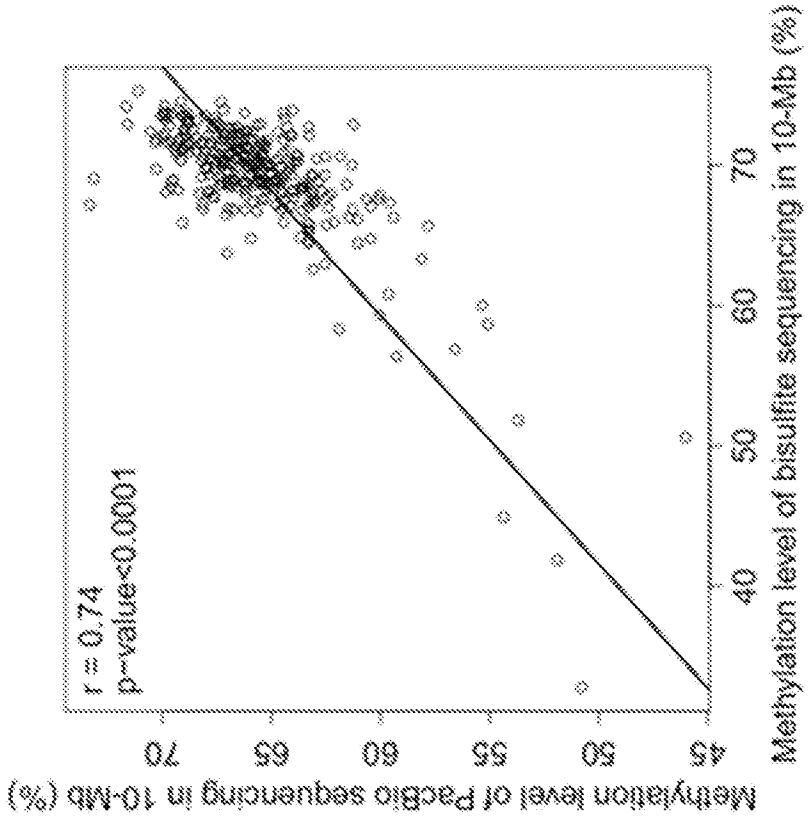


FIG. 52B

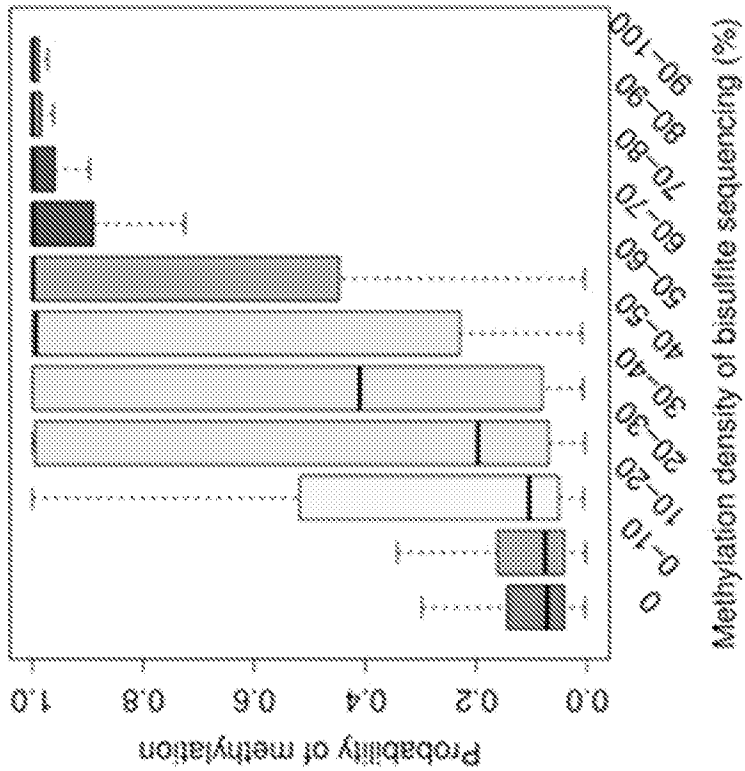


FIG. 52A

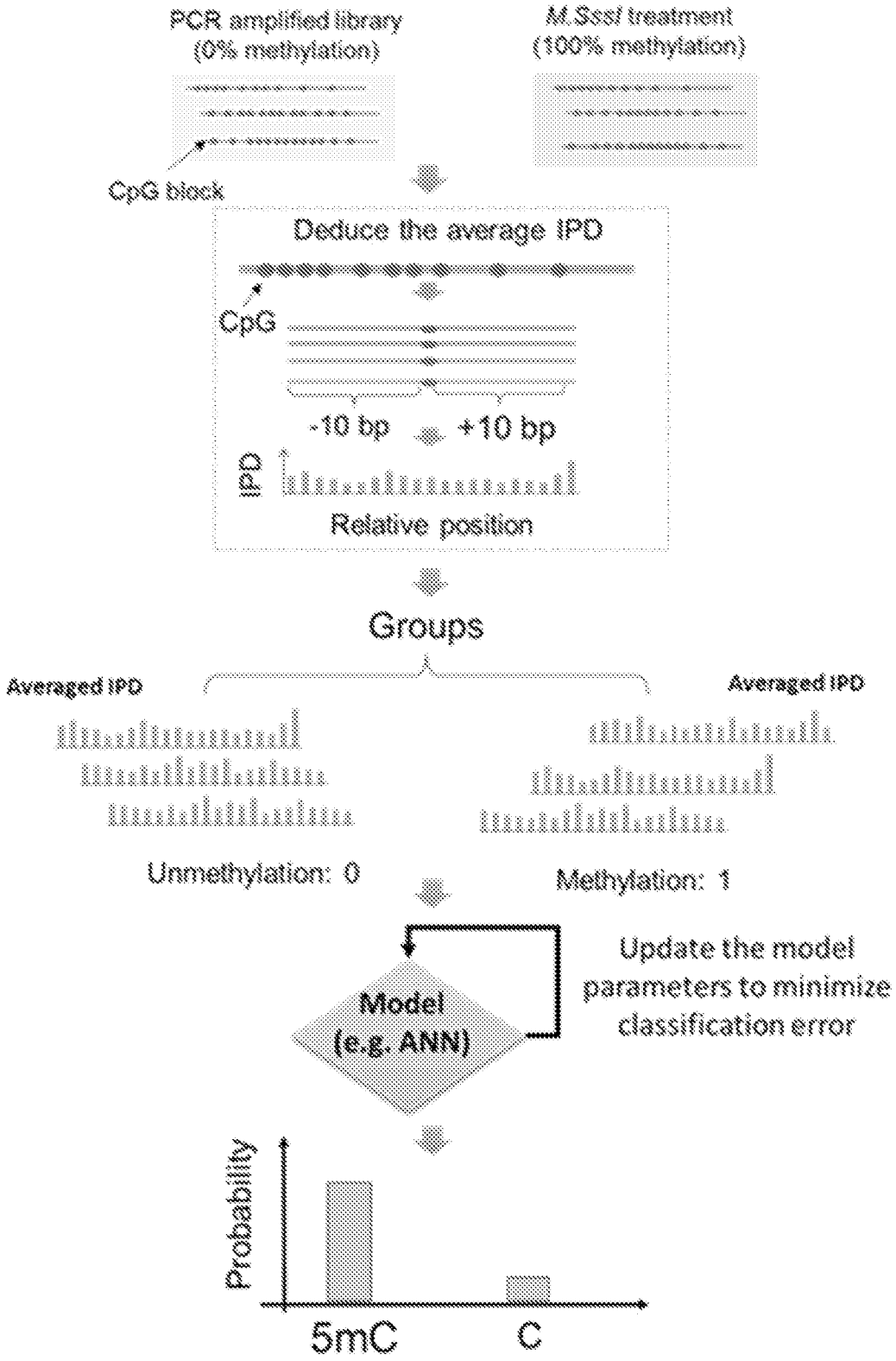

PCR amplified library
(0% methylation)
M.SssI treatment
(100% methylation)
CpG block
Deduce the average IPD
CpG
-10 bp
+10 bp
IPD
Relative position
Groups
Averaged IPD
Averaged IPD
Unmethylation: 0
Methylation: 1
Model
(e.g. ANN)
Update the model
parameters to minimize
classification error
Probability
5mC
C

FIG. 54

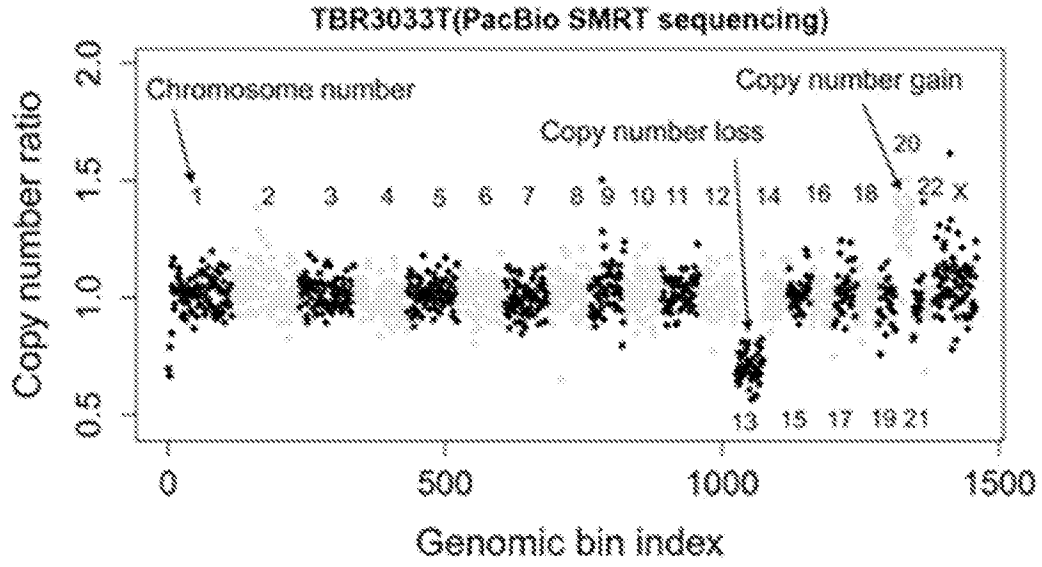


FIG. 56A

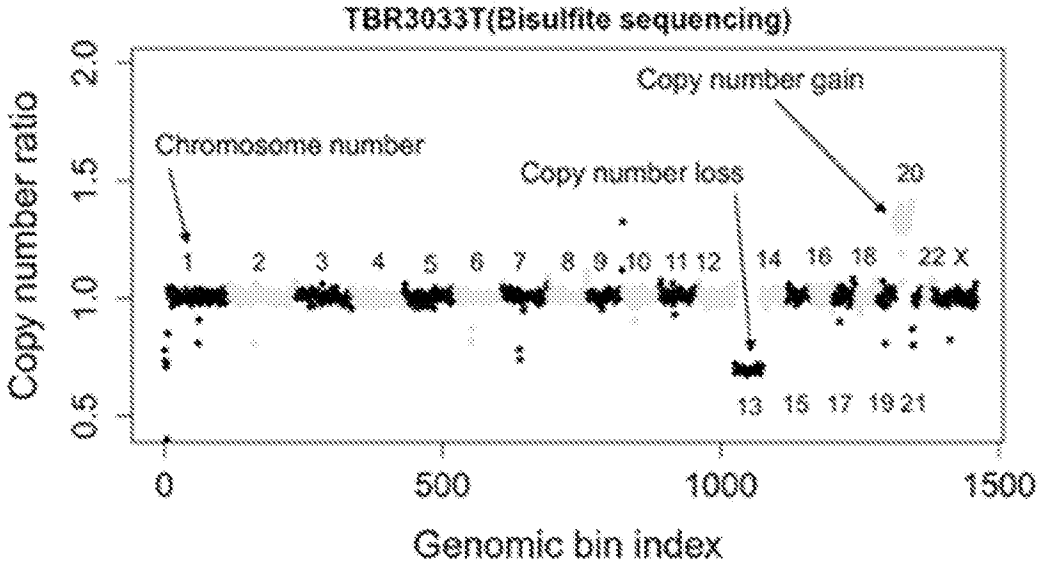


FIG. 56B

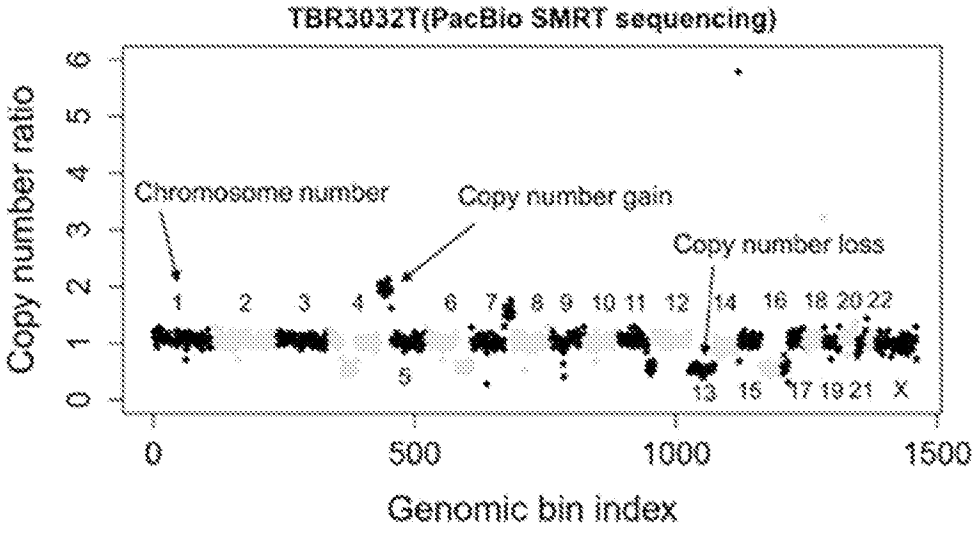


FIG. 57A

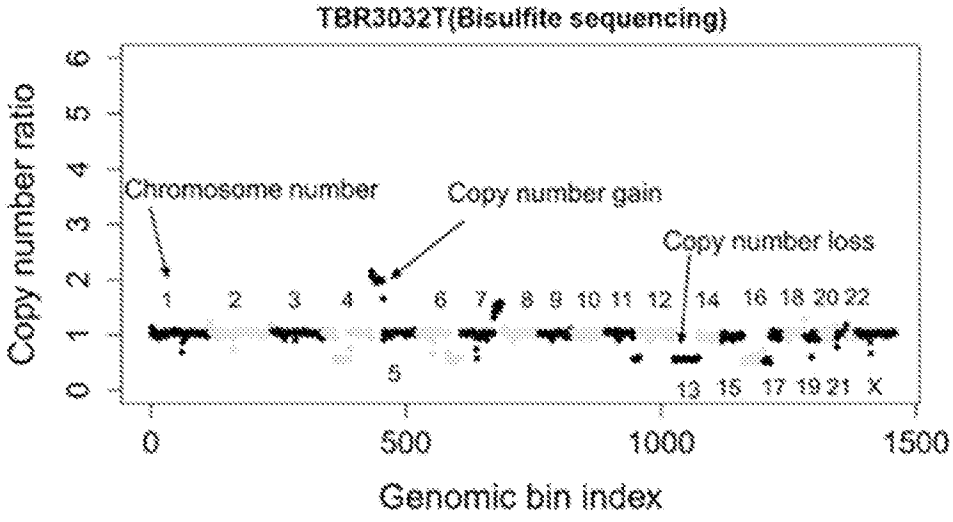


FIG. 57B

Plasma DNA
Circulating fetal DNA
Maternal DNA
PacBio sequencing
Bisulfite sequencing
Reference tissue methylome
T cell
B cell
Neutrophil
Liver
Lung
Colon
Placenta
Plasma DNA tissue mapping
% DNA contribution
100
50
0
Neutrophil
B cell
T cell
Liver
Lung
Colon
Placenta

FIG. 58

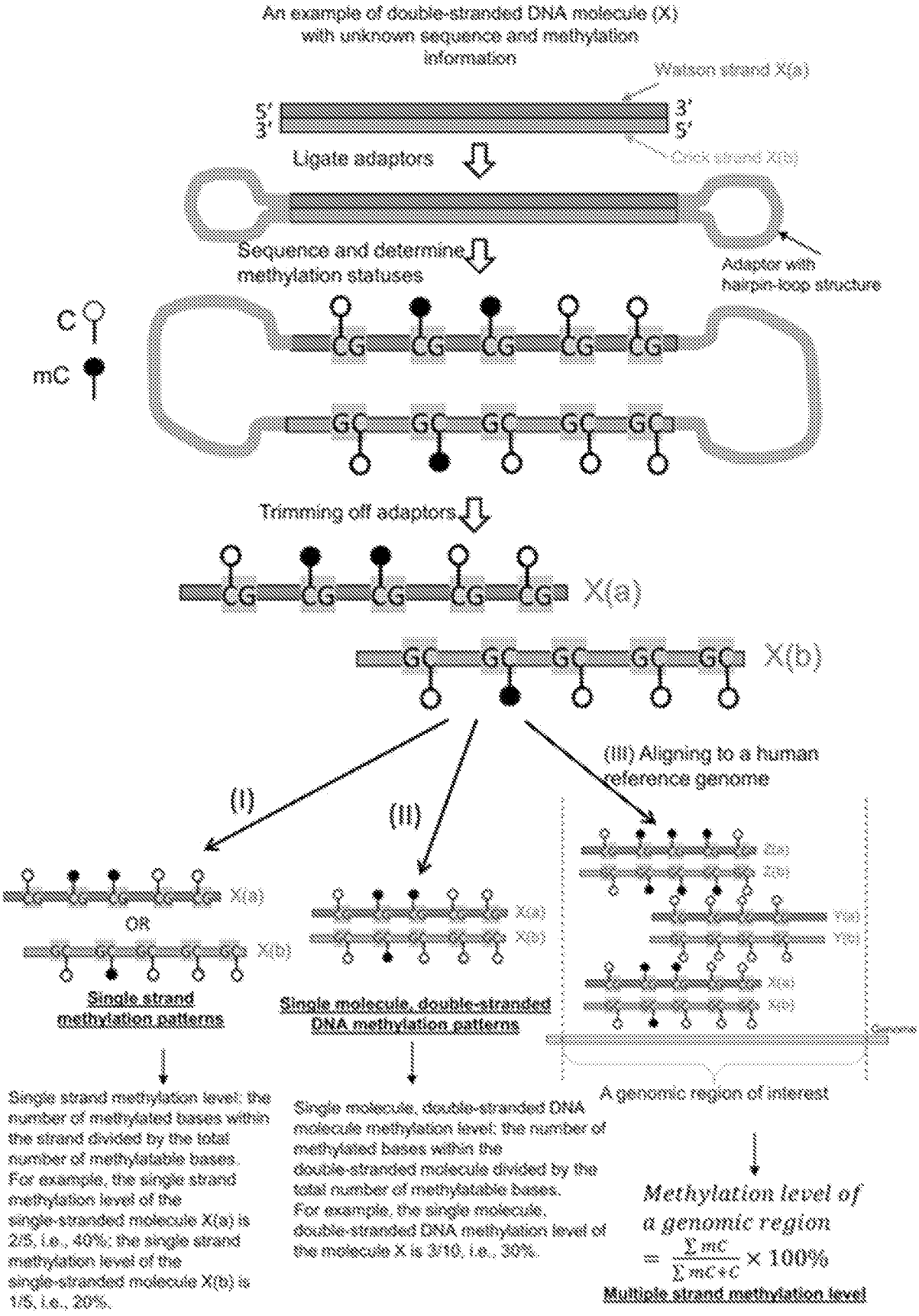

An example of double-stranded DNA molecule (X) with unknown sequence and methylation information
Watson strand X(a)
5'
3'
3'
5'
Crick strand X(b)
Ligate adaptors
Sequence and determine methylation statuses
Adaptor with hairpin-loop structure
C
mC
CG
GC
Trimming off adaptors
X(a)
X(b)
(I)
(II)
(III) Aligning to a human reference genome
X(a)
OR
X(b)
Single strand methylation patterns
Single molecule, double-stranded DNA methylation patterns
A genomic region of interest
Single strand methylation level: the number of methylated bases within the strand divided by the total number of methylatable bases. For example, the single strand methylation level of the single-stranded molecule X(a) is 2/5, i.e., 40%; the single strand methylation level of the single-stranded molecule X(b) is 1/5, i.e., 20%.
Single molecule, double-stranded DNA molecule methylation level: the number of methylated bases within the double-stranded molecule divided by the total number of methylatable bases. For example, the single molecule, double-stranded DNA methylation level of the molecule X is 3/10, i.e., 30%.
Methylation level of a genomic region
$= \frac{\sum mC}{\sum mC+C} \times 100\%$
Multiple strand methylation level

FIG. 61

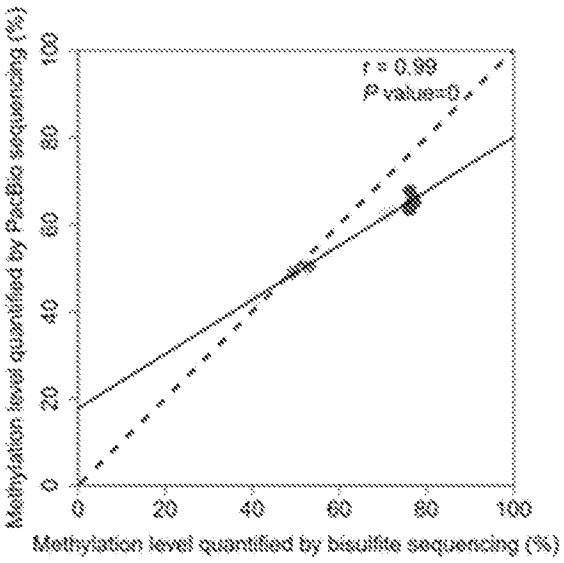


FIG. 63A

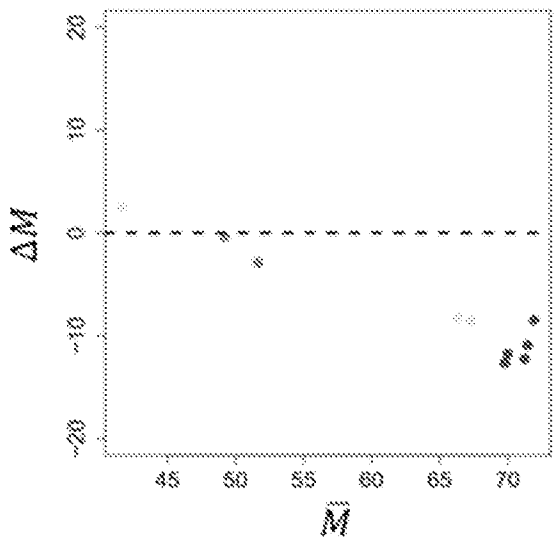


FIG. 63B

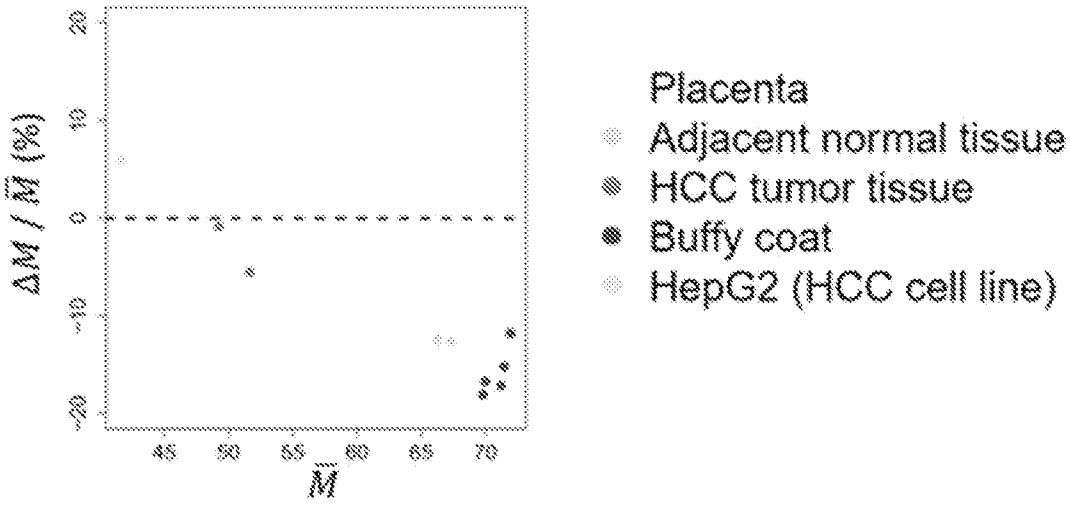


FIG. 63C

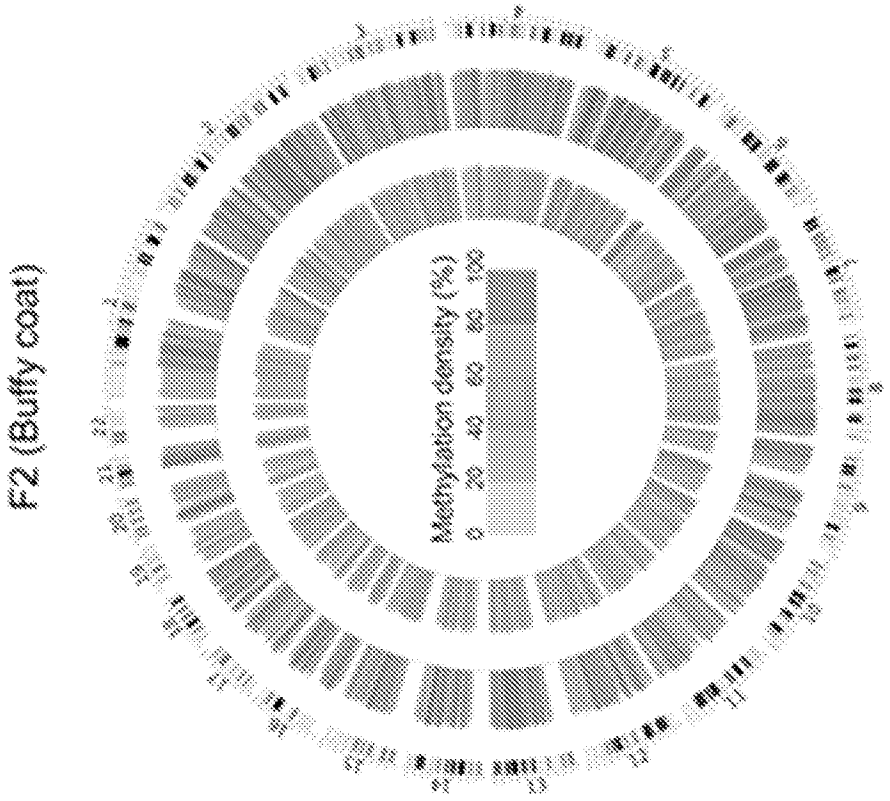


FIG. 64B

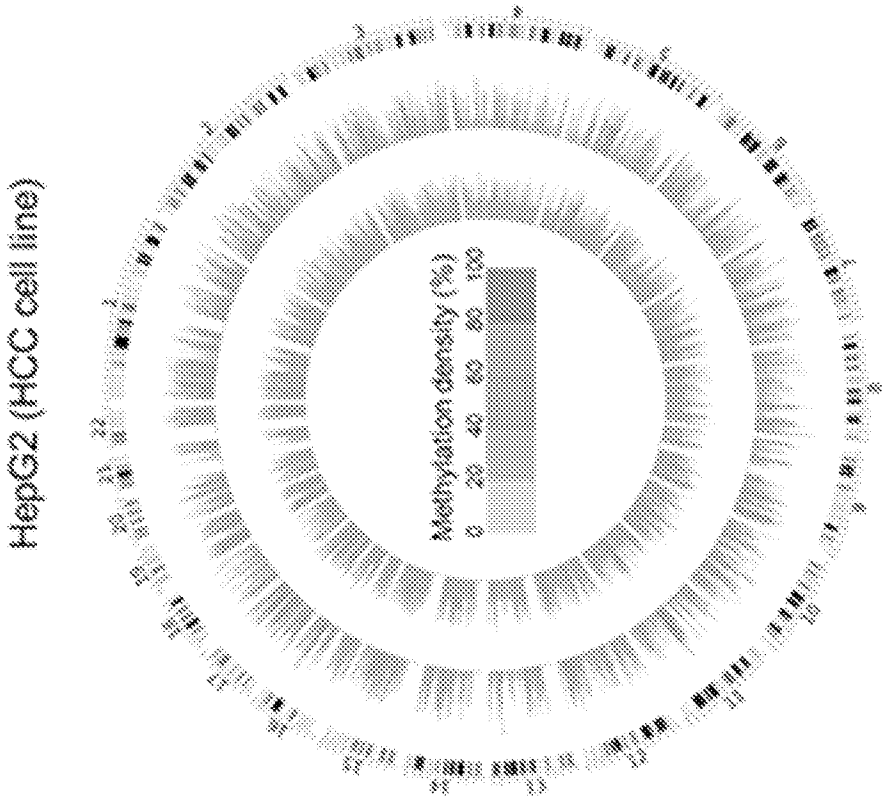


FIG. 64A

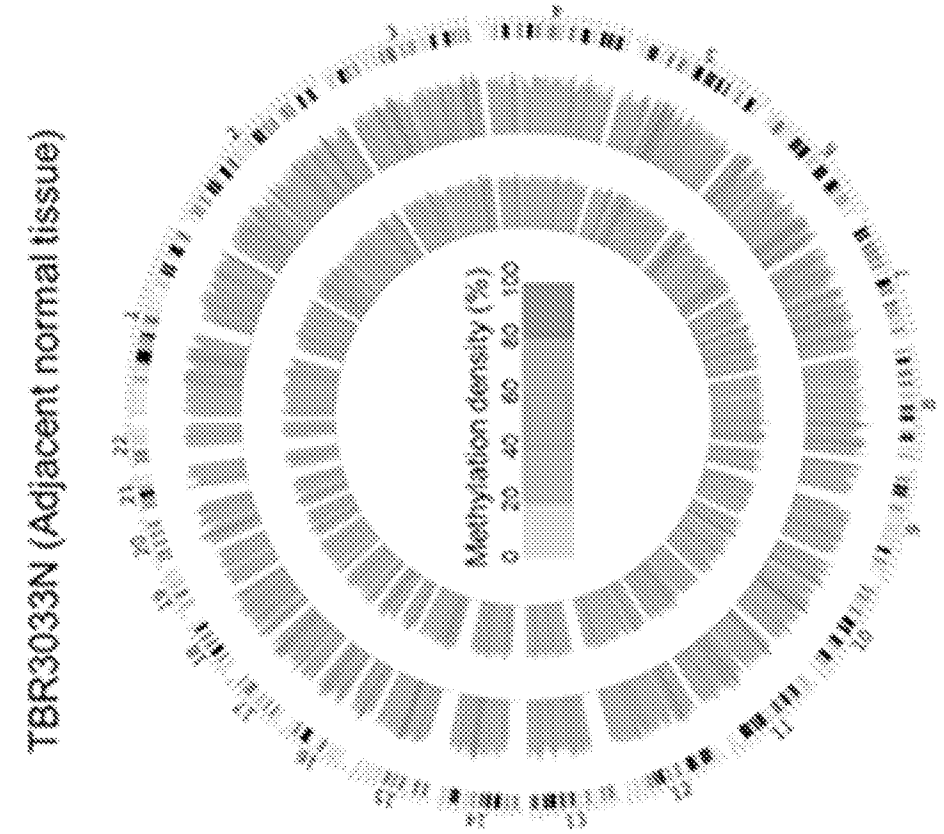


FIG. 67B

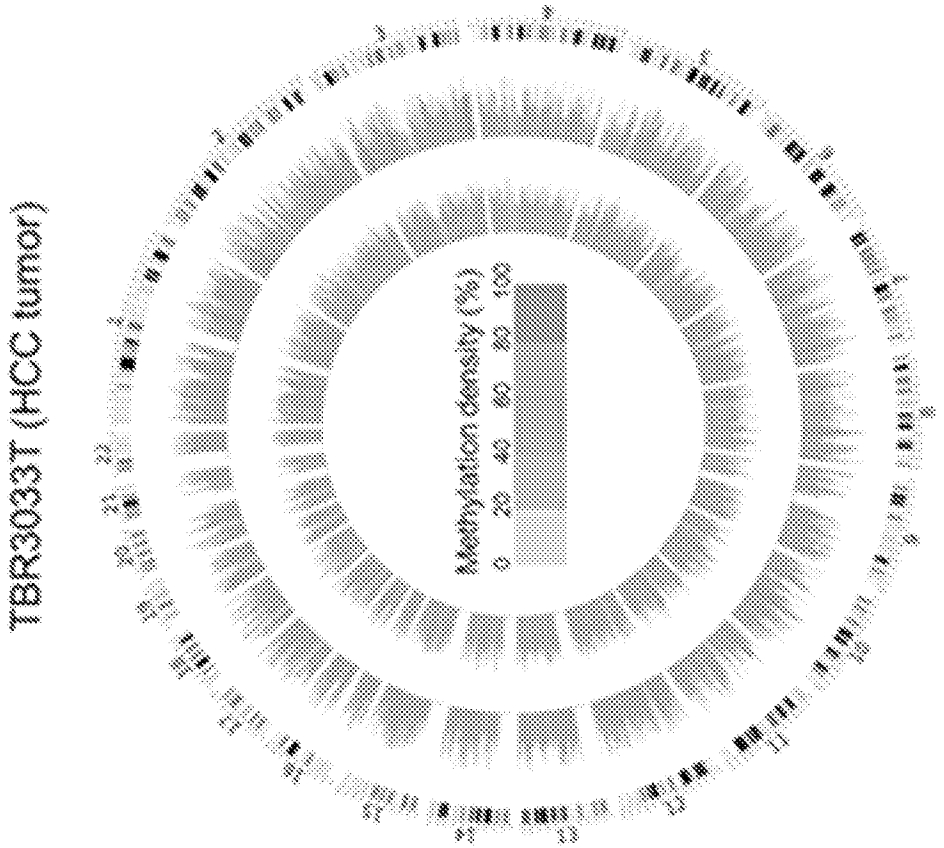


FIG. 67A

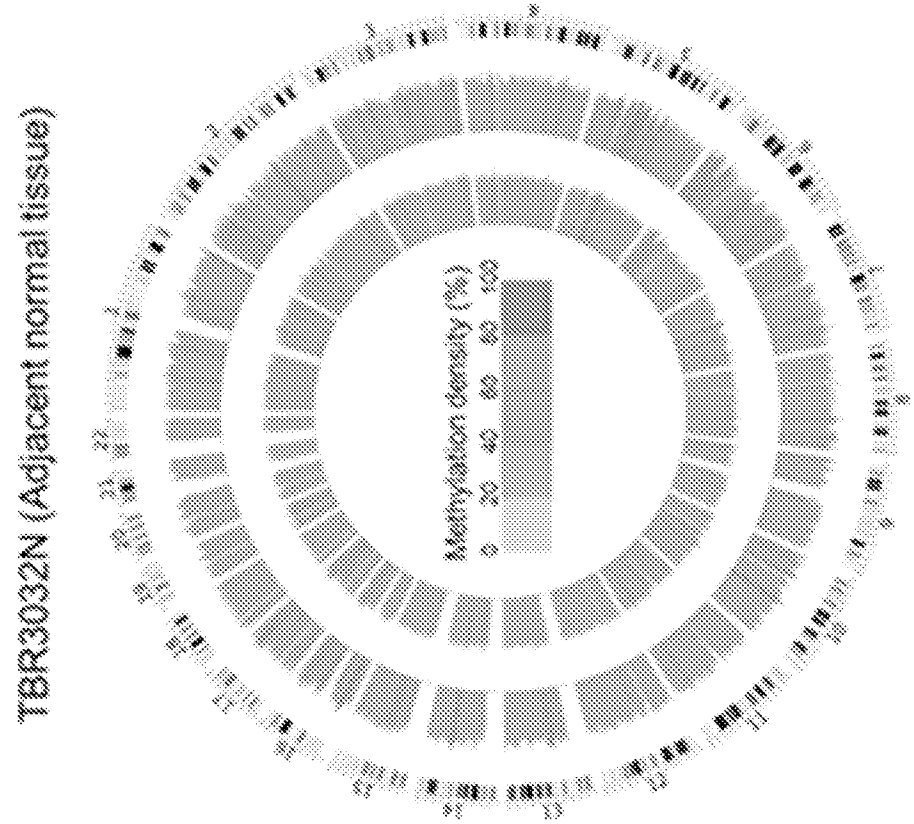


FIG. 70B

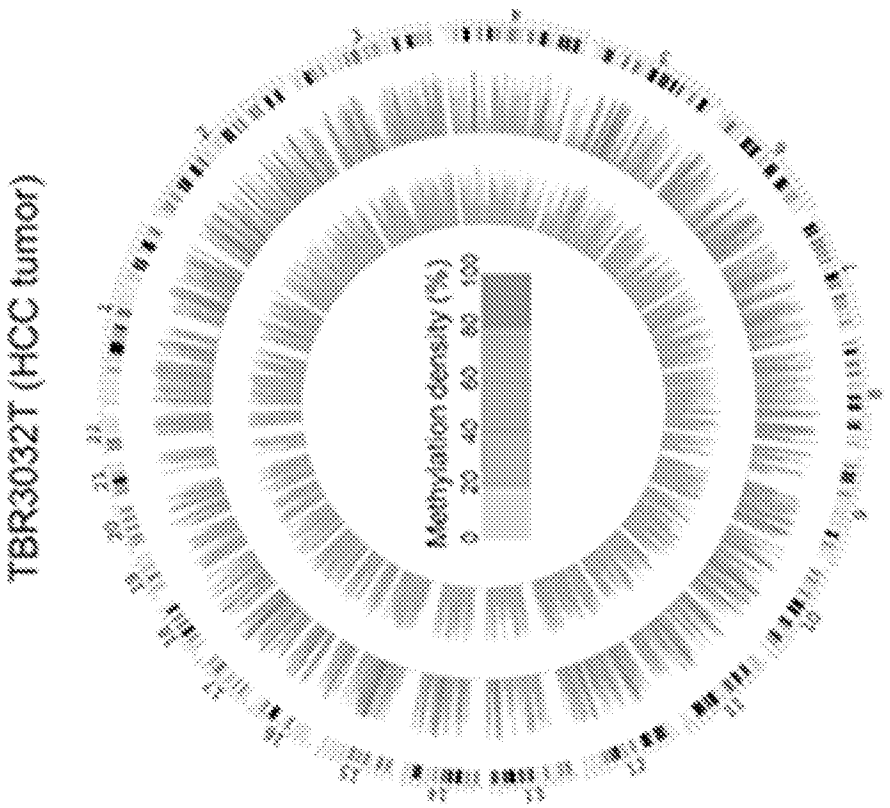


FIG. 70A

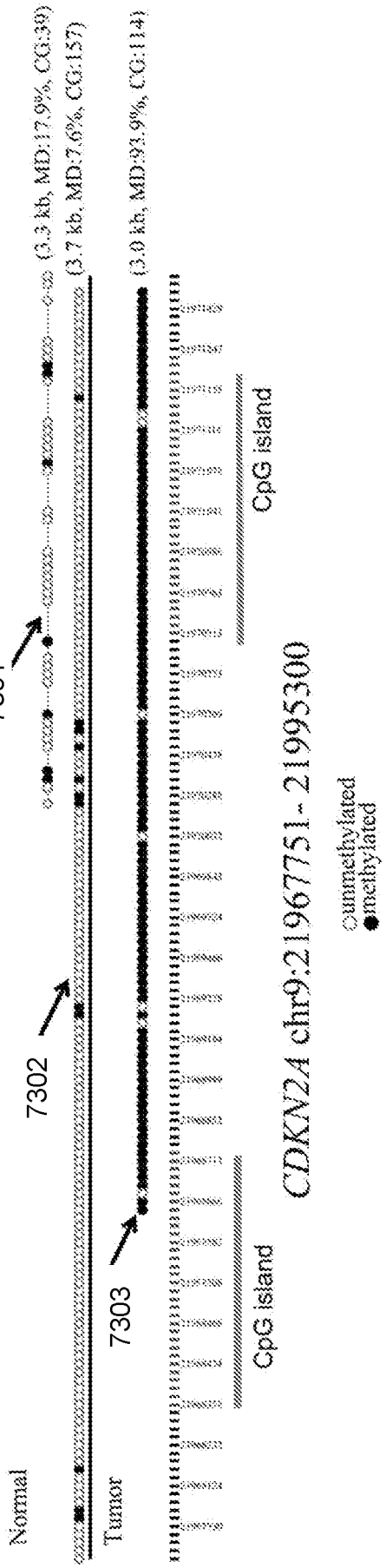

TBR3032
7301
7302
7303
(3.3 kb, MD:17.9%, CG:39)
(3.7 kb, MD:7.6%, CG:157)
(3.0 kb, MD:93.9%, CG:114)
CpG island
CpG island
CDKN2A chr9:21967751-21995300
unmethylated
methylated
Normal
Tumor

FIG. 73

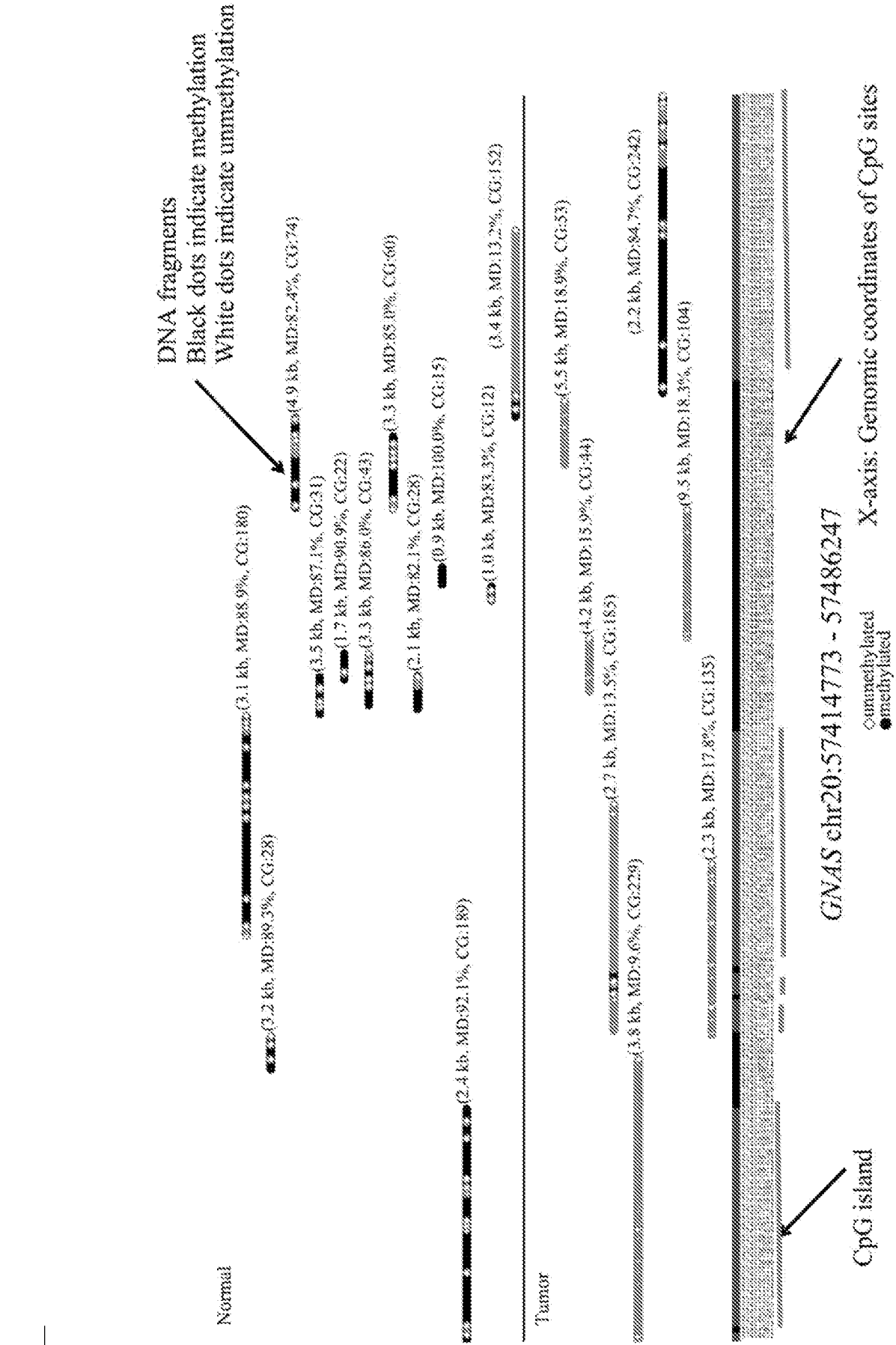

DNA fragments
Black dots indicate methylation
White dots indicate unmethylation
Normal
Tumor
(3.1 kb, MD:88.9%, CG:180)
(3.2 kb, MD:89.3%, CG:28)
(4.9 kb, MD:82.4%, CG:74)
(3.5 kb, MD:87.1%, CG:31)
(1.7 kb, MD:90.9%, CG:22)
(3.3 kb, MD:86.0%, CG:43)
(3.3 kb, MD:85.0%, CG:60)
(2.1 kb, MD:82.1%, CG:28)
(0.9 kb, MD:100.0%, CG:15)
(2.4 kb, MD:92.1%, CG:189)
(1.0 kb, MD:83.3%, CG:12)
(3.4 kb, MD:13.2%, CG:152)
(5.5 kb, MD:16.9%, CG:53)
(4.2 kb, MD:15.9%, CG:44)
(2.7 kb, MD:13.5%, CG:185)
(2.2 kb, MD:84.7%, CG:342)
(4.8 kb, MD:9.6%, CG:229)
(9.5 kb, MD:18.3%, CG:104)
(2.3 kb, MD:17.8%, CG:135)
GNAS chr20:57414773 - 57486247
unmethylated
methylated
X-axis: Genomic coordinates of CpG sites
CpG island

FIG. 74A

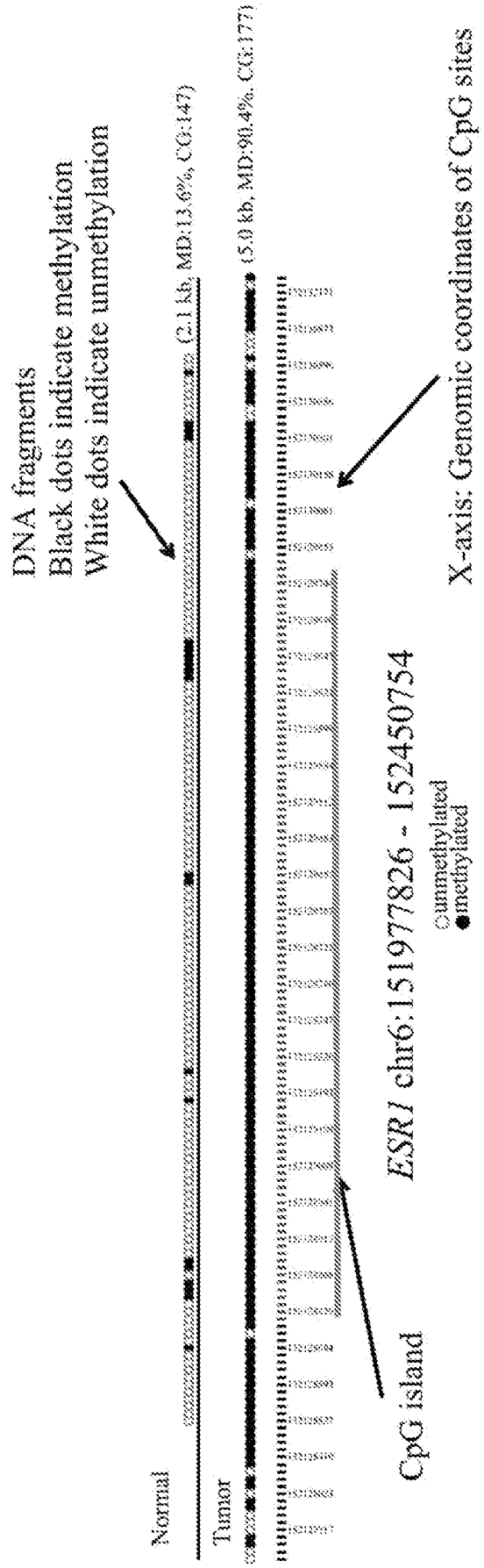

DNA fragments
Black dots indicate methylation
White dots indicate unmethylation
(2.1 kb, MD:13.6%, CG:147)
(5.0 kb, MD:90.4%, CG:177)
Normal
Tumor
CpG island
ESR1 chr6:151977826 - 152450754
○unmethylated
●methylated
X-axis: Genomic coordinates of CpG sites

FIG. 74B

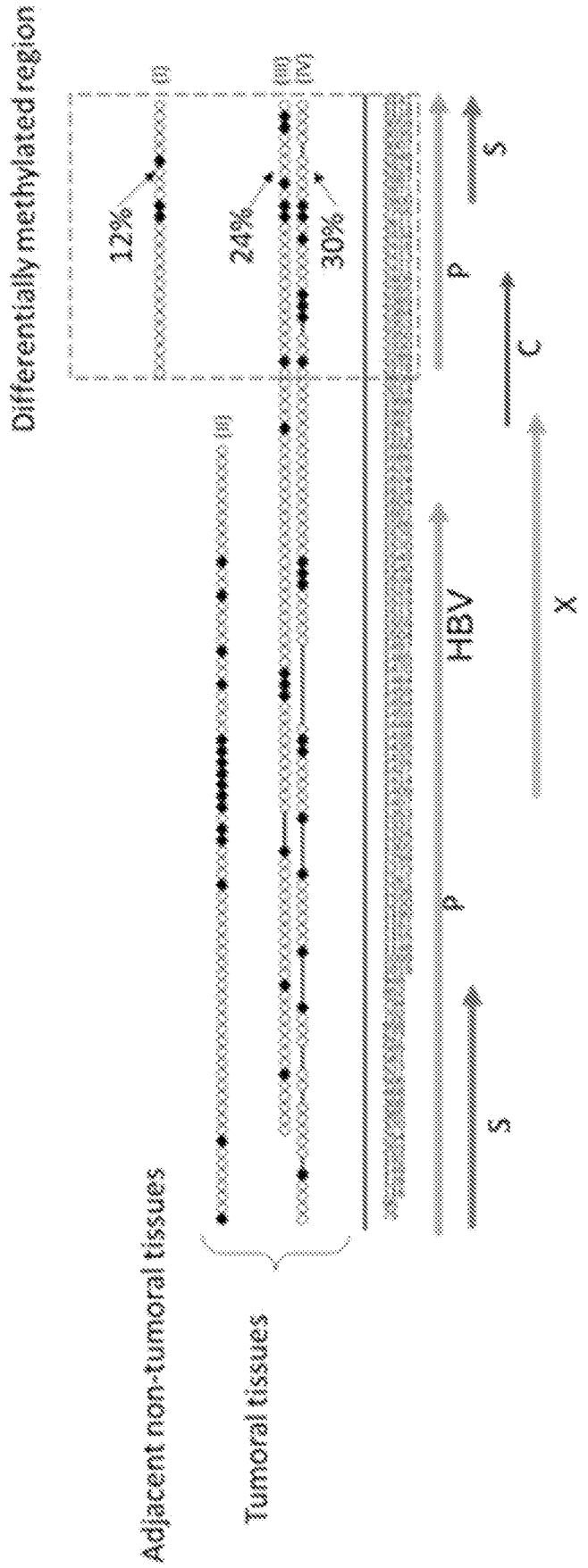
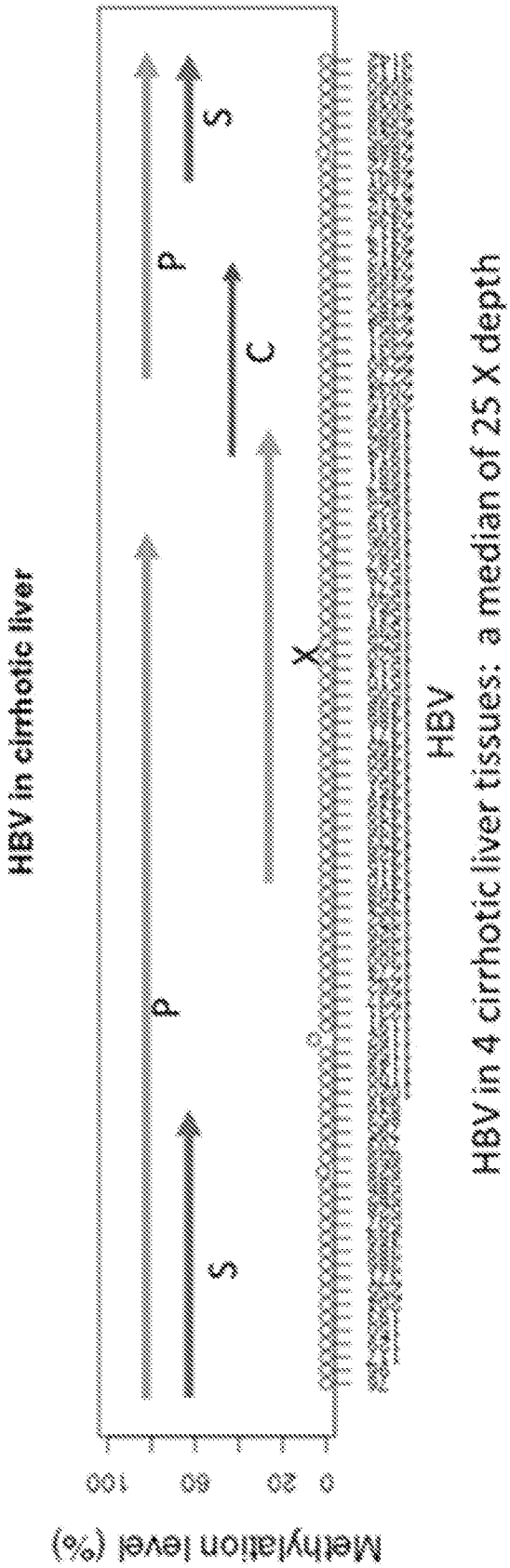
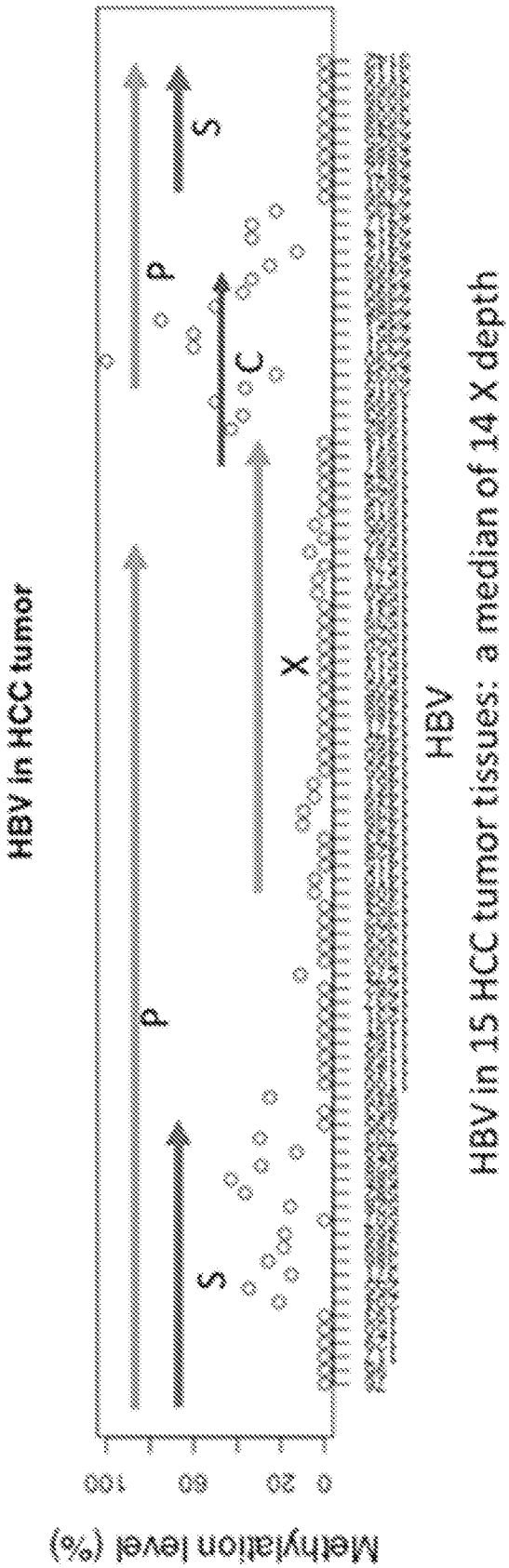

Differentially methylated region
12%
24%
30%
(I)
(III)
(IV)
(II)
Adjacent non-tumoral tissues
Tumoral tissues
HBV
P
S
X
C
P
S

FIG. 75



HBV in 4 cirrhotic liver tissues: a median of 25 X depth

FIG. 76A



HBV in 15 HCC tumor tissues: a median of 14 X depth

FIG. 76B

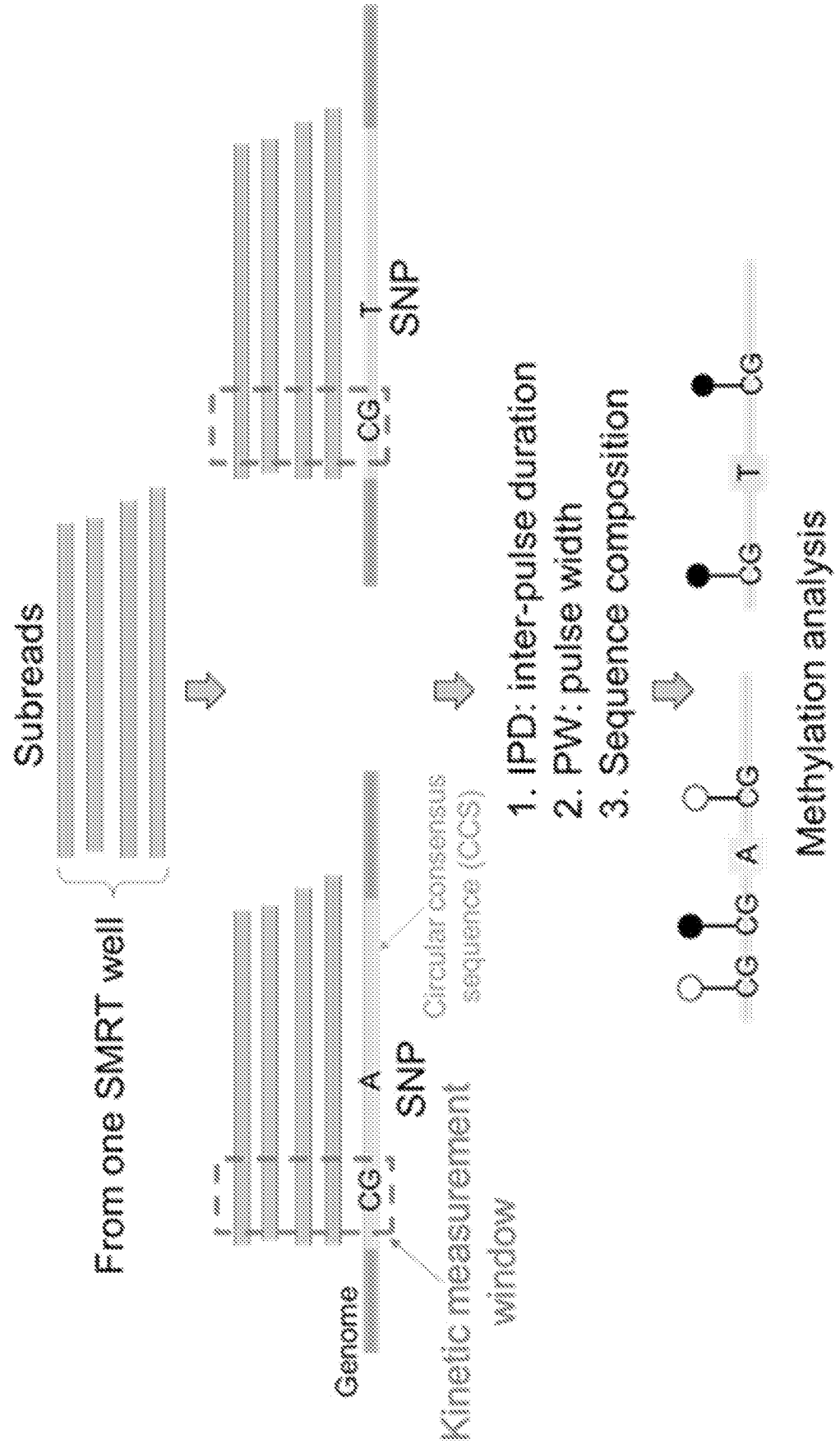

Subreads
From one SMRT well
Genome
CG
A
SNP
Circular consensus sequence (CCS)
Kinetic measurement window
CG
T
SNP
1. IPD: inter-pulse duration
2. PW: pulse width
3. Sequence composition
CG
CG
A
CG
CG
T
CG
Methylation analysis

FIG. 77

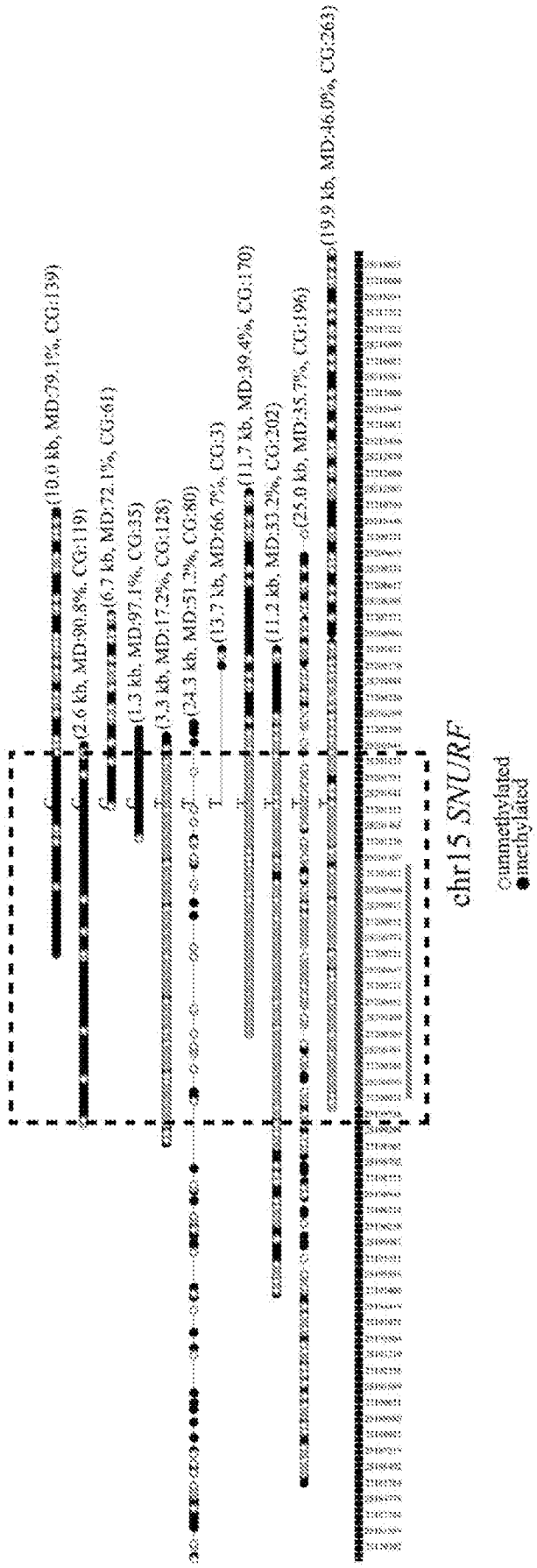

(10.0 kb, MD:79.1%, CG:139)
(2.6 kb, MD:90.8%, CG:119)
(6.7 kb, MD:72.1%, CG:61)
(1.3 kb, MD:97.1%, CG:35)
(3.3 kb, MD:17.2%, CG:128)
(24.3 kb, MD:51.2%, CG:80)
(13.7 kb, MD:66.7%, CG:3)
(11.7 kb, MD:39.4%, CG:170)
(11.2 kb, MD:33.2%, CG:202)
(25.0 kb, MD:35.7%, CG:196)
(19.9 kb, MD:46.6%, CG:263)
chr15 SNURF
unmethylated
methylated

FIG. 79A

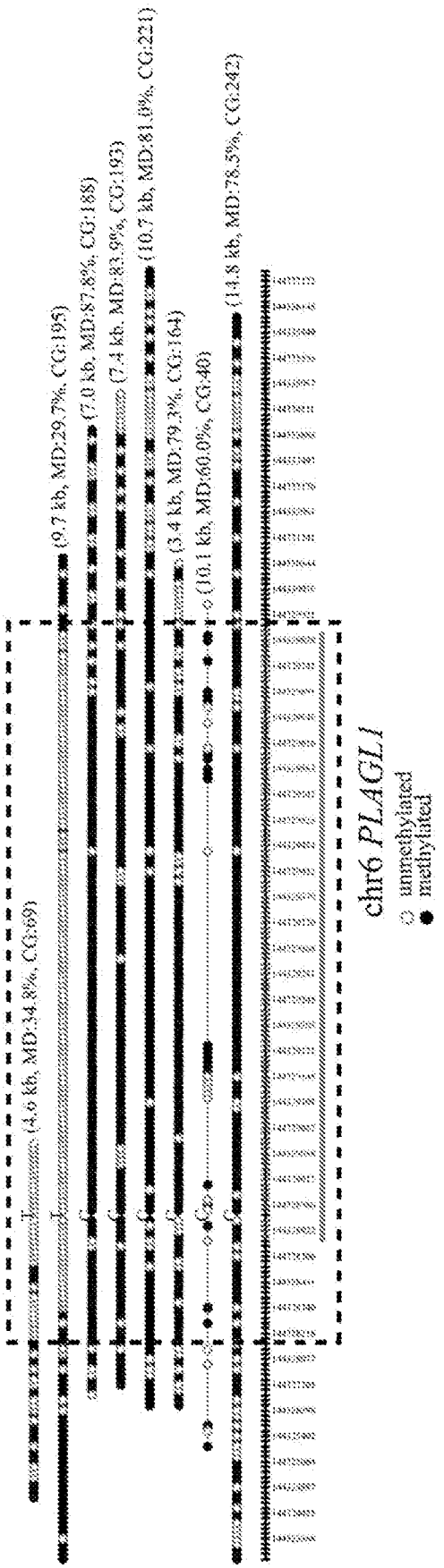

(4.6 kb, MD:34.8%, CG:69)
(9.7 kb, MD:29.7%, CG:195)
(7.0 kb, MD:87.8%, CG:188)
(7.4 kb, MD:83.9%, CG:193)
(10.7 kb, MD:81.0%, CG:221)
(3.4 kb, MD:79.3%, CG:164)
(10.1 kb, MD:60.0%, CG:40)
(14.8 kb, MD:78.5%, CG:242)
chr6 PLAGL1
unmethylated
methylated

FIG. 79B

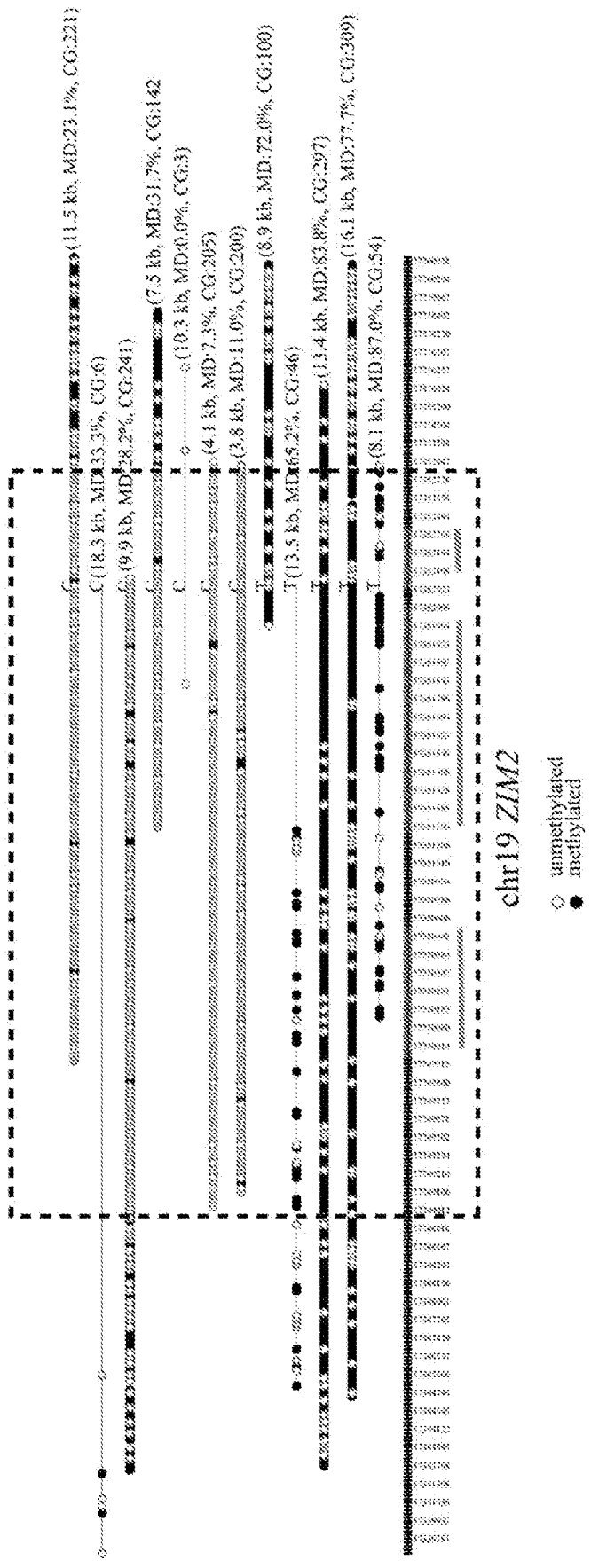

(11.5 kb, MD:23.1%, CG:221)
C (18.3 kb, MD:33.3%, CG:6)
(9.9 kb, MD:28.2%, CG:241)
(7.5 kb, MD:31.7%, CG:142
(10.3 kb, MD:0.0%, CG:3)
(4.1 kb, MD:7.3%, CG:205)
(3.8 kb, MD:11.0%, CG:200)
(8.9 kb, MD:72.0%, CG:100)
T (13.5 kb, MD:65.2%, CG:46)
(13.4 kb, MD:83.8%, CG:297)
(16.1 kb, MD:77.7%, CG:309)
(8.1 kb, MD:87.0%, CG:54)
chr19 ZIM2
unmethylated
methylated

FIG. 79D

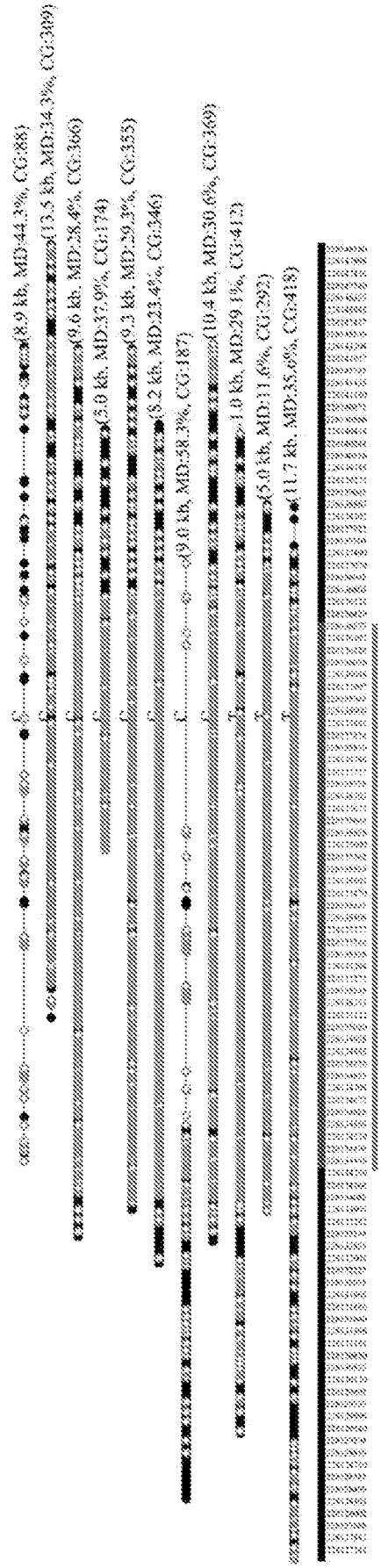

(8.9 kb, MD:44.3%, CG:288)
(13.5 kb, MD:34.3%, CG:309)
(9.6 kb, MD:28.4%, CG:306)
(3.0 kb, MD:37.9%, CG:174)
(9.3 kb, MD:29.3%, CG:355)
(8.2 kb, MD:23.4%, CG:346)
(9.0 kb, MD:58.3%, CG:187)
(10.4 kb, MD:30.6%, CG:369)
(10 kb, MD:29.1%, CG:412)
(5.0 kb, MD:11.6%, CG:292)
(11.7 kb, MD:35.6%, CG:418)
chr1:228125898-228145899
unmethylated
methylated

FIG. 80D

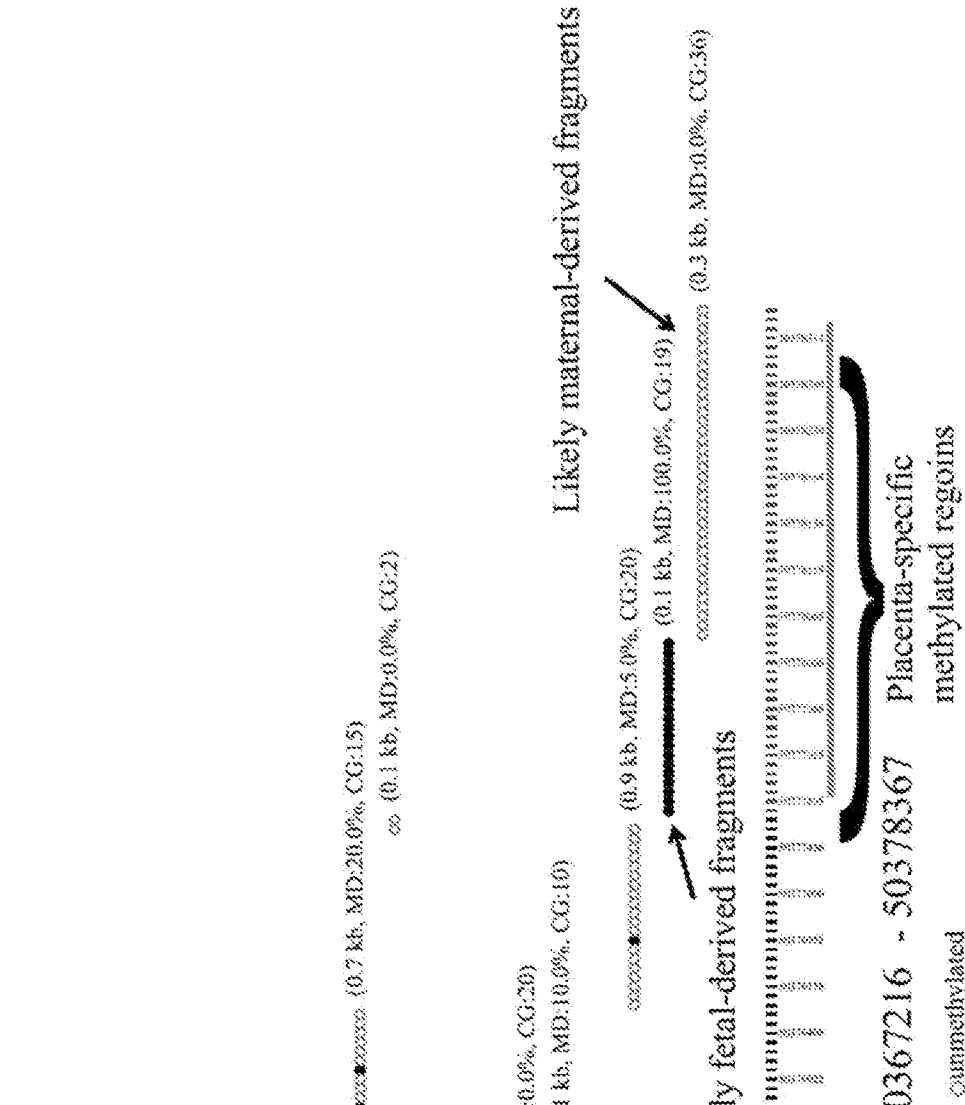

(0.7 kb, MD:20.0%, CG:15)
(0.1 kb, MD:0.0%, CG:2)
(0.1 kb, MD:100.0%, CG:4)
(0.1 kb, MD:71.4%, CG:7)
(1.0 kb, MD:76.9%, CG:13)
(0.1 kb, MD:0.0%, CG:20)
(0.1 kb, MD:10.0%, CG:10)
(0.1 kb, MD:100.0%, CG:3)
(0.9 kb, MD:5.0%, CG:20)
(0.1 kb, MD:100.0%, CG:19)
(0.3 kb, MD:0.0%, CG:36)
Likely maternal-derived fragments
Likely fetal-derived fragments
(0.1 kb, MD:100.0%, CG:3)
RASSF1A chr3:50367216 - 50378367
Placenta-specific methylated regoins
unmethylated
methylated

FIG. 82

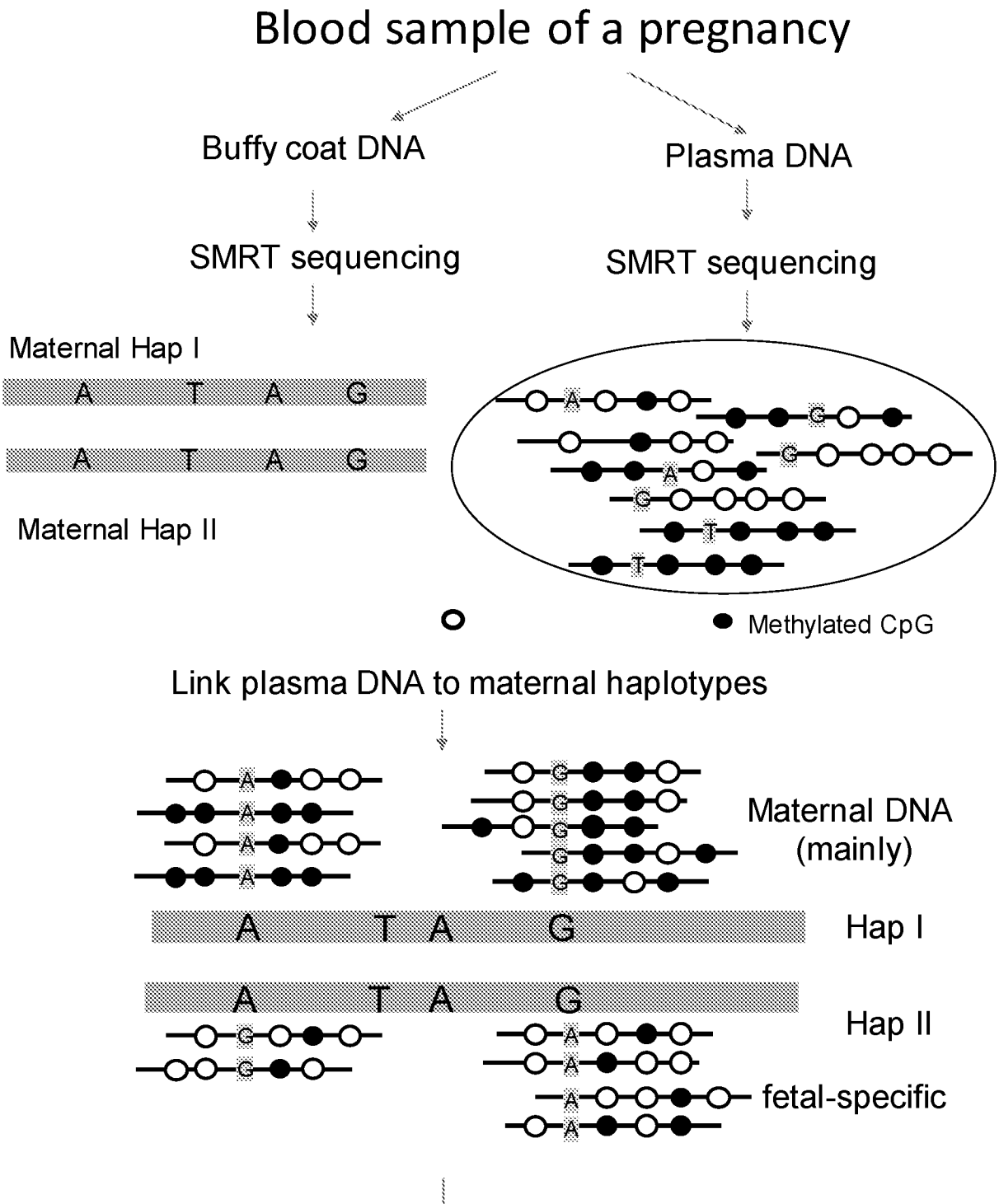

Blood sample of a pregnancy
Buffy coat DNA
Plasma DNA
SMRT sequencing
SMRT sequencing
Maternal Hap I
A T A G
A T A G
Maternal Hap II
Methylated CpG
Link plasma DNA to maternal haplotypes
Maternal DNA (mainly)
A T A G
Hap I
A T A G
Hap II
fetal-specific
The methylation level of fetal-specific DNA would be lower compared with that of maternal fragments

FIG. 83

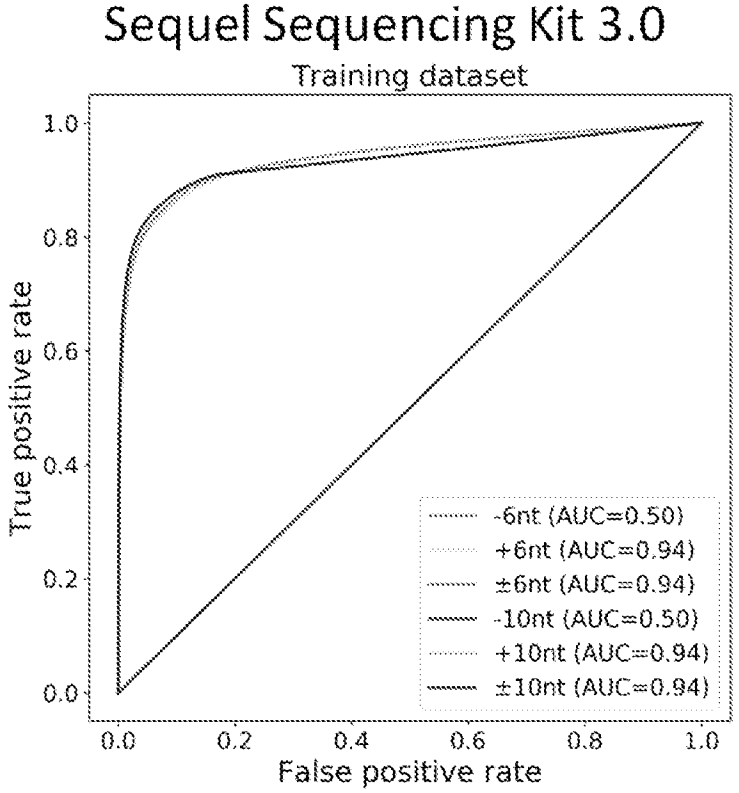


FIG. 84A

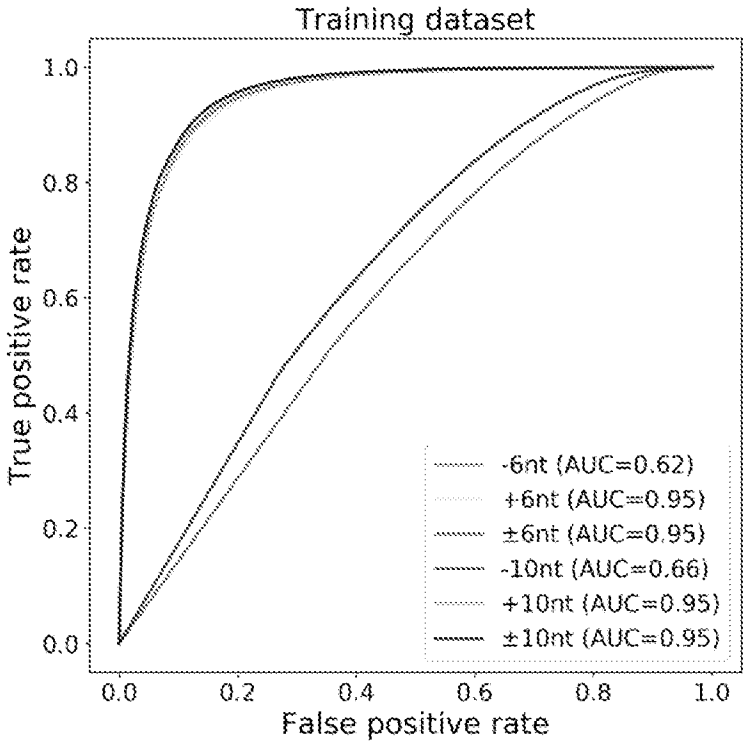


FIG. 84B

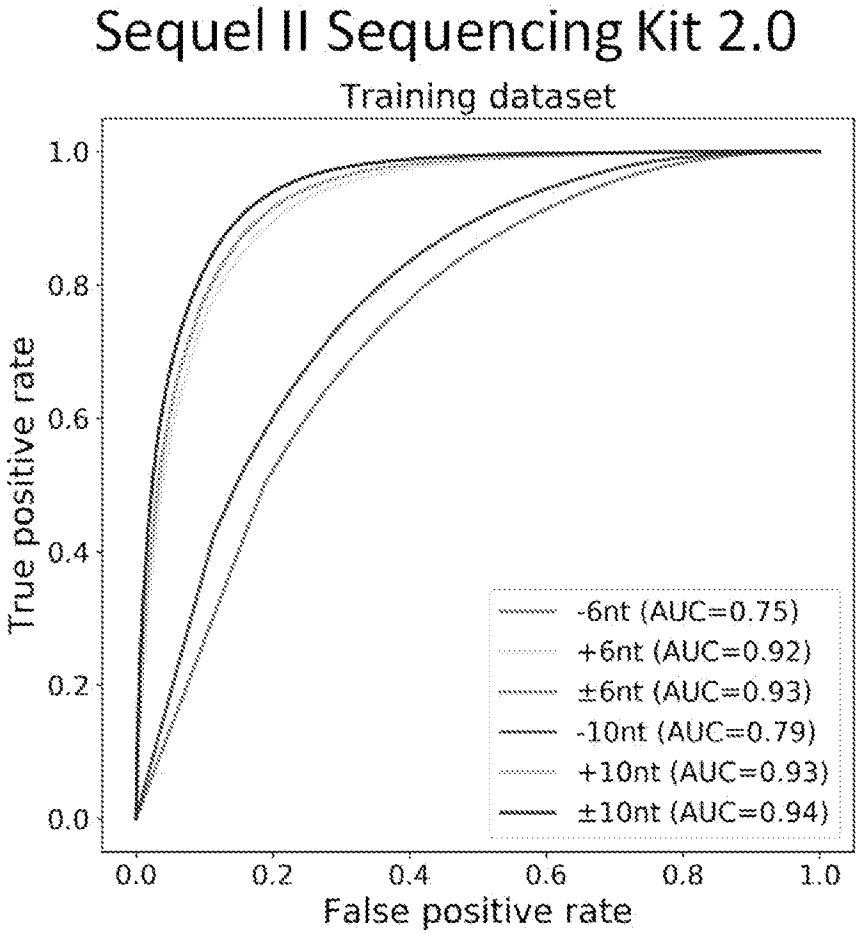

Sequel II Sequencing Kit 2.0
Training dataset
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
-6nt (AUC=0.75)
+6nt (AUC=0.92)
±6nt (AUC=0.93)
-10nt (AUC=0.79)
+10nt (AUC=0.93)
±10nt (AUC=0.94)

FIG. 84C

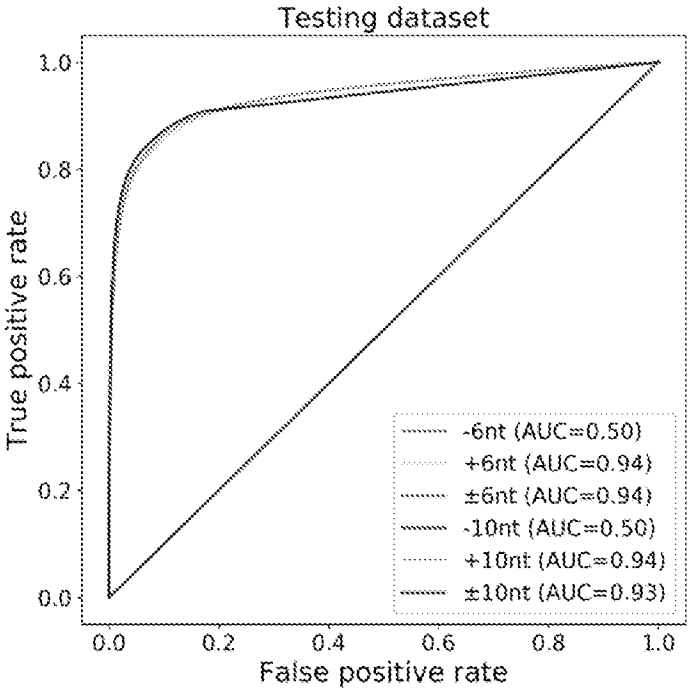


FIG. 85A

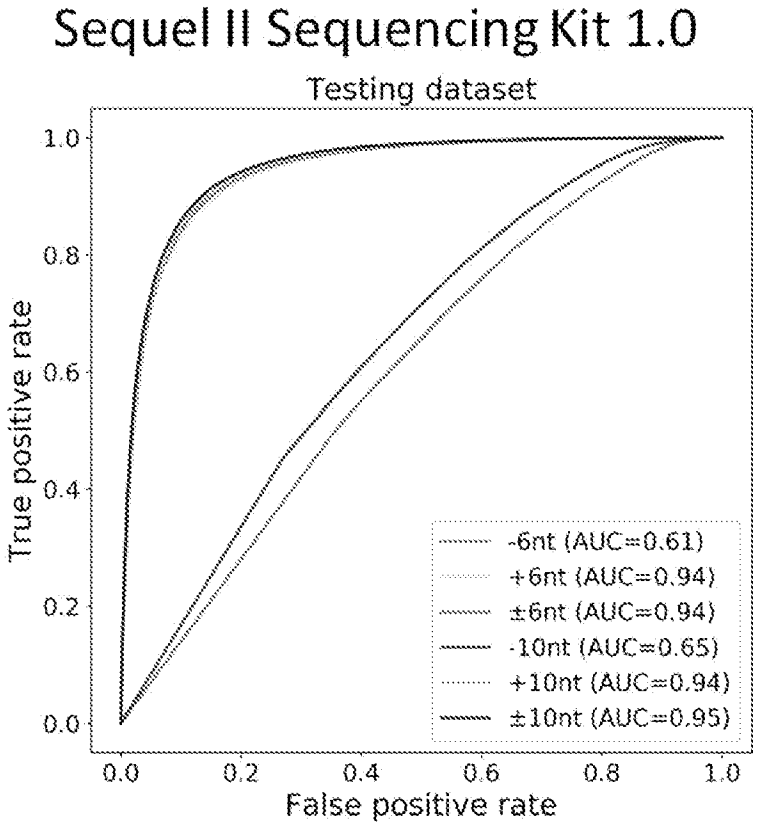


FIG. 85B

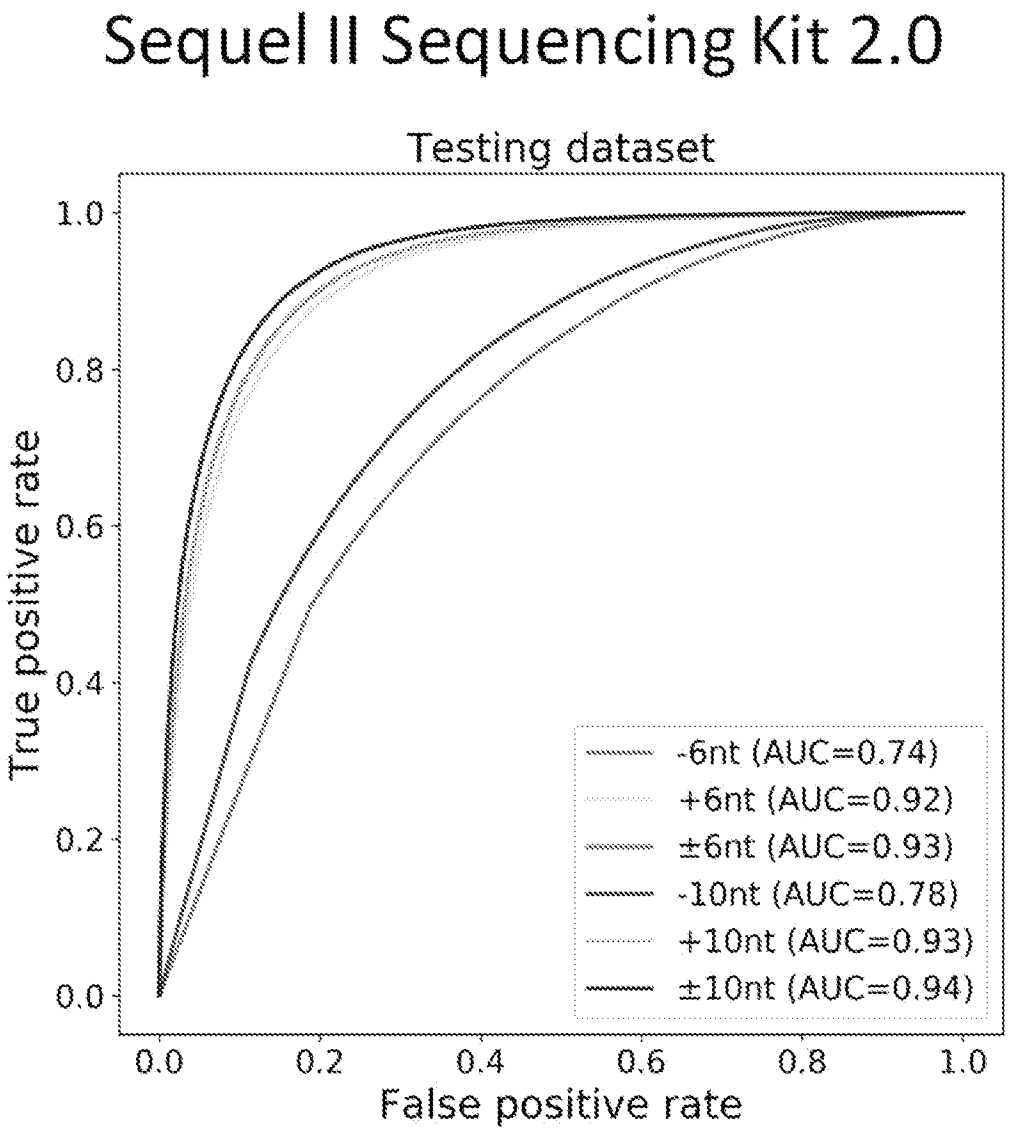

Sequel II Sequencing Kit 2.0
Testing dataset
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
-6nt (AUC=0.74)
+6nt (AUC=0.92)
±6nt (AUC=0.93)
-10nt (AUC=0.78)
+10nt (AUC=0.93)
±10nt (AUC=0.94)

FIG. 85C

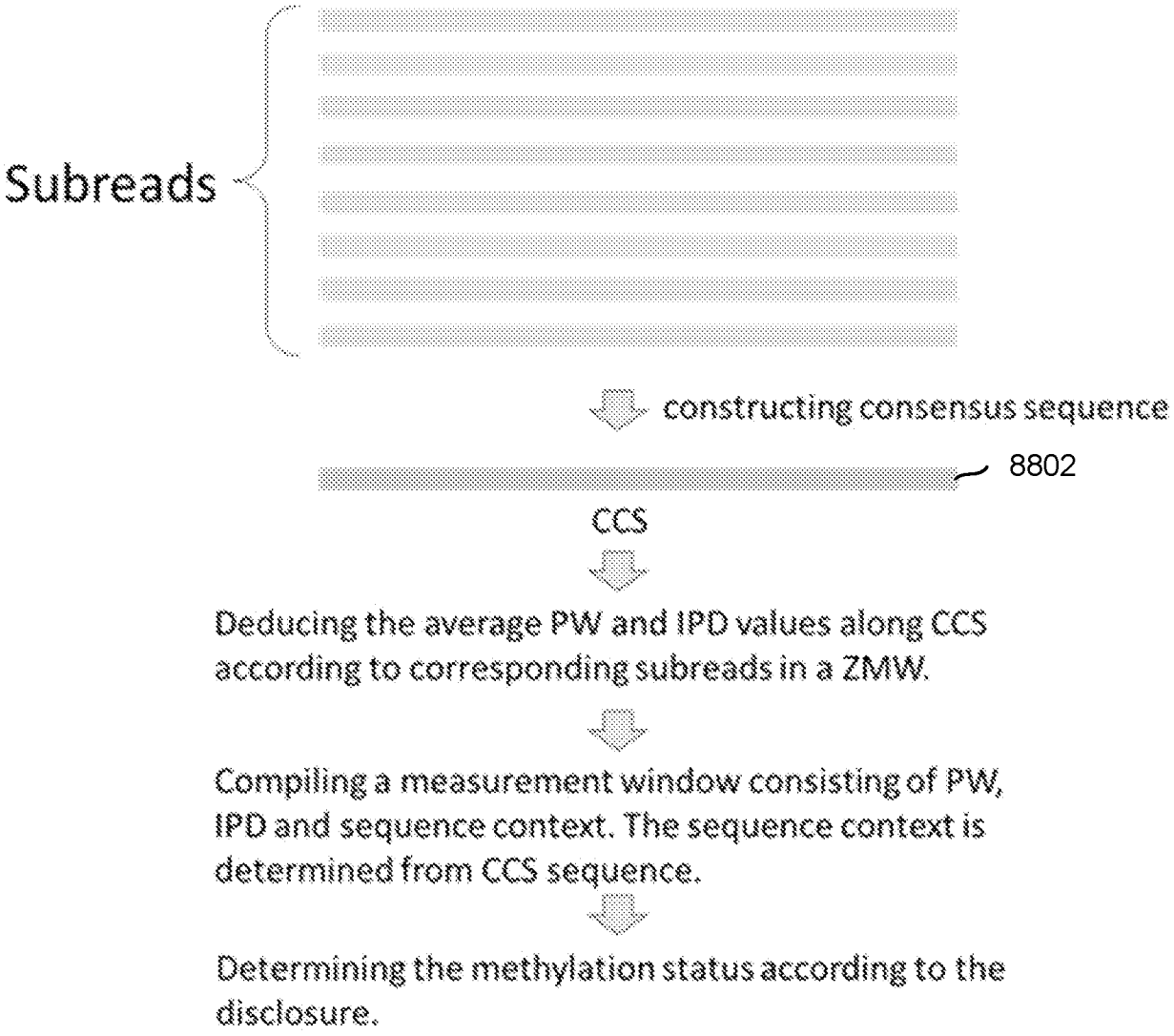

Subreads
constructing consensus sequence
8802
CCS
Deducing the average PW and IPD values along CCS according to corresponding subreads in a ZMW.
Compiling a measurement window consisting of PW, IPD and sequence context. The sequence context is determined from CCS sequence.
Determining the methylation status according to the disclosure.

FIG. 88

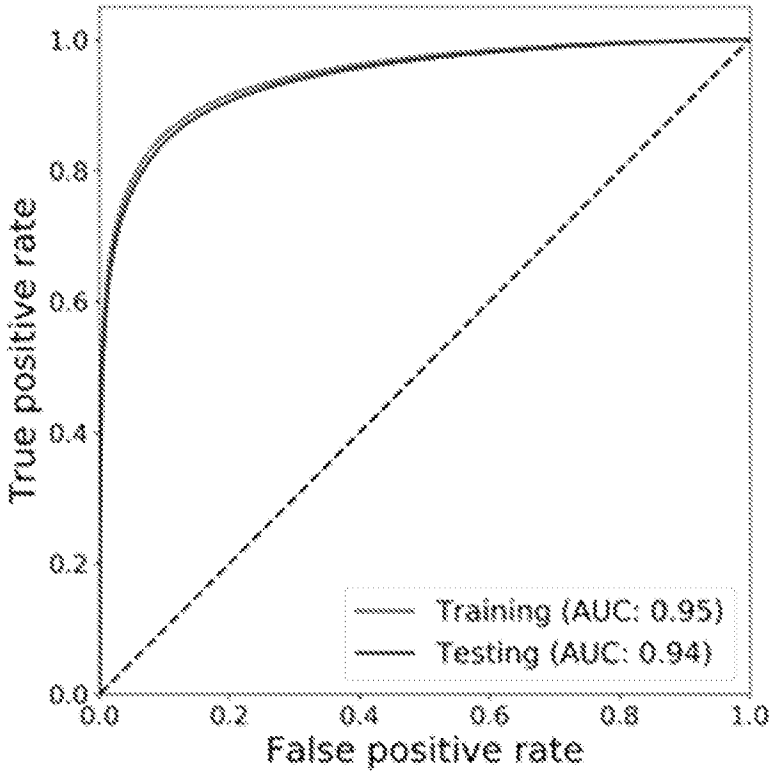

True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Training (AUC: 0.95)
Testing (AUC: 0.94)

FIG. 89

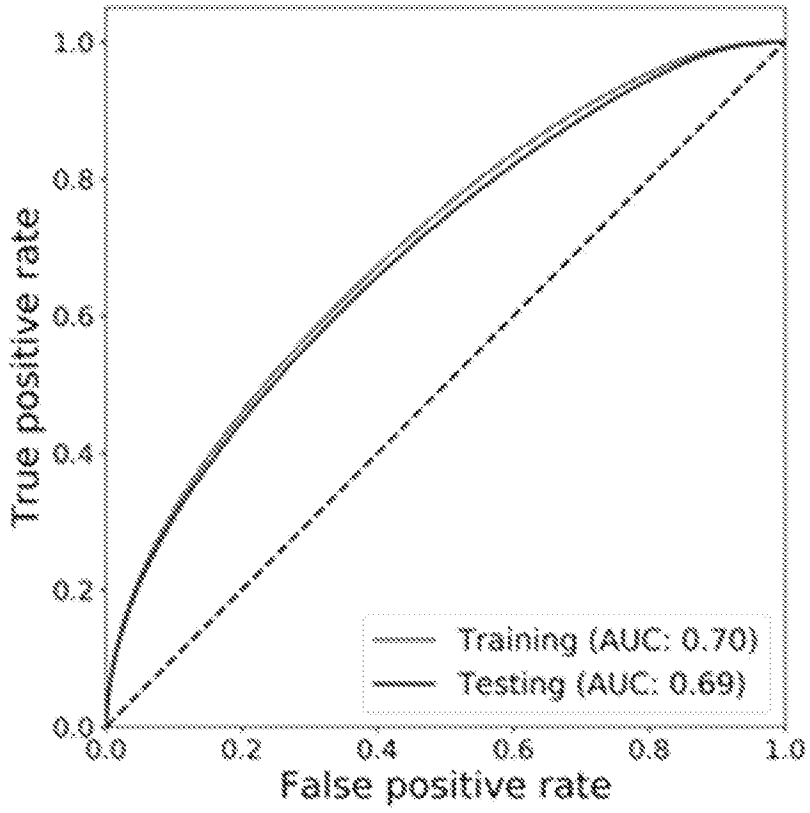

True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Training (AUC: 0.70)
Testing (AUC: 0.69)

FIG. 90

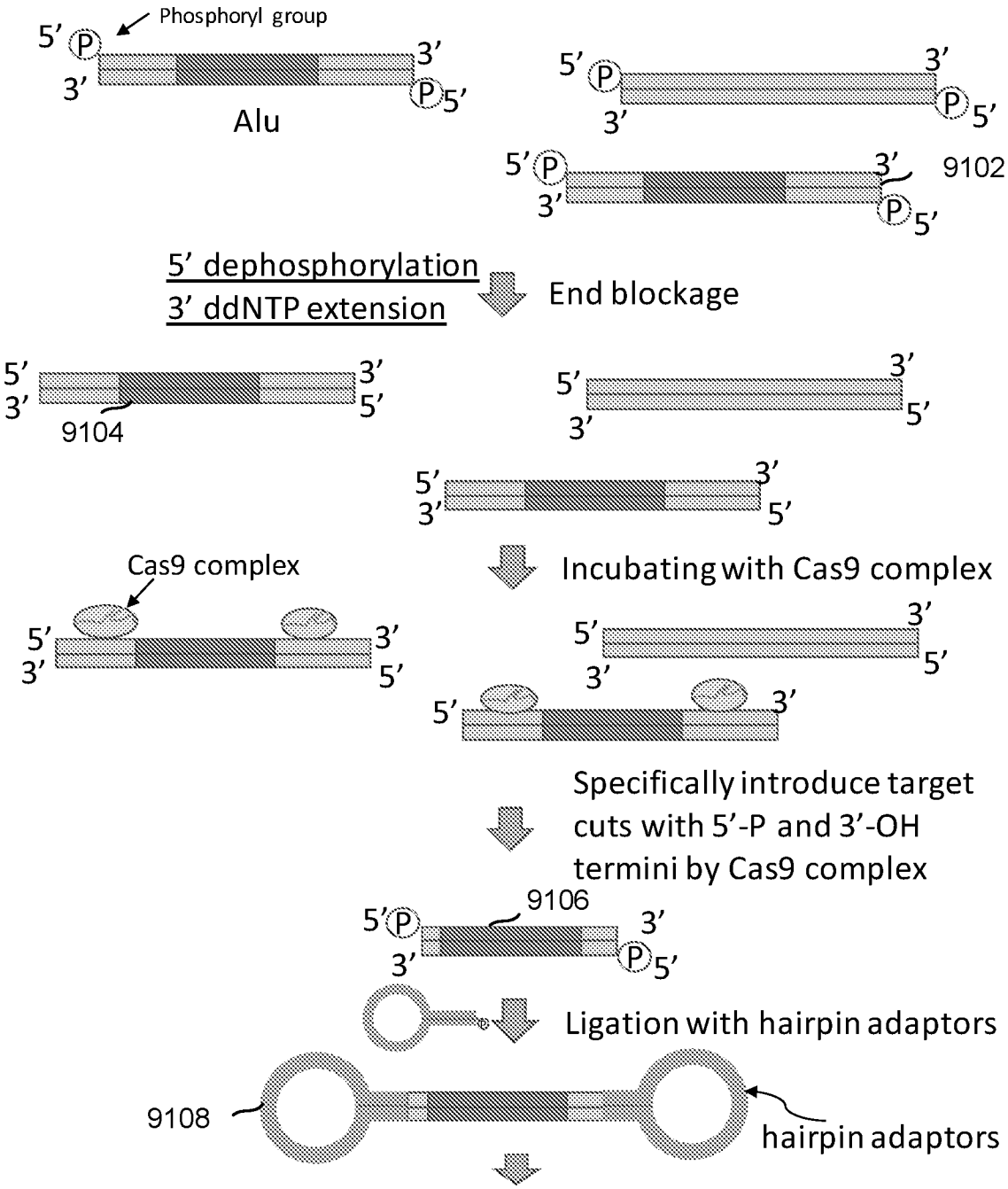

Phosphoryl group
5'
3'
Alu
9102
5' dephosphorylation
3' ddNTP extension
End blockage
9104
Cas9 complex
Incubating with Cas9 complex
Specifically introduce target cuts with 5'-P and 3'-OH termini by Cas9 complex
9106
Ligation with hairpin adaptors
9108
hairpin adaptors
Remove the adaptors, linear DNA and off-target DNA using exonuclease (e.g. exonuclease III and VII)
Single molecule, real-time sequencing
Base modification analysis using IPD, PW and sequence context according to the disclosure

FIG. 91

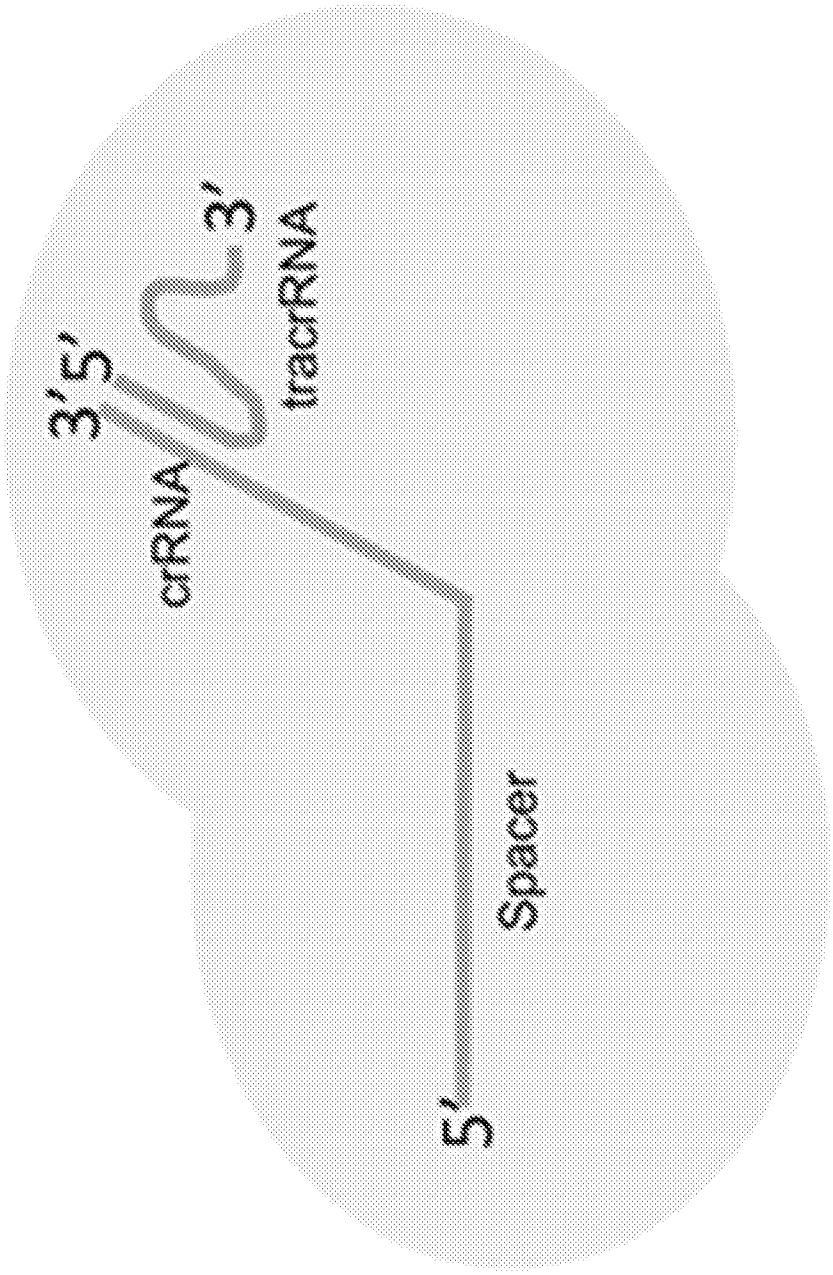


FIG. 92

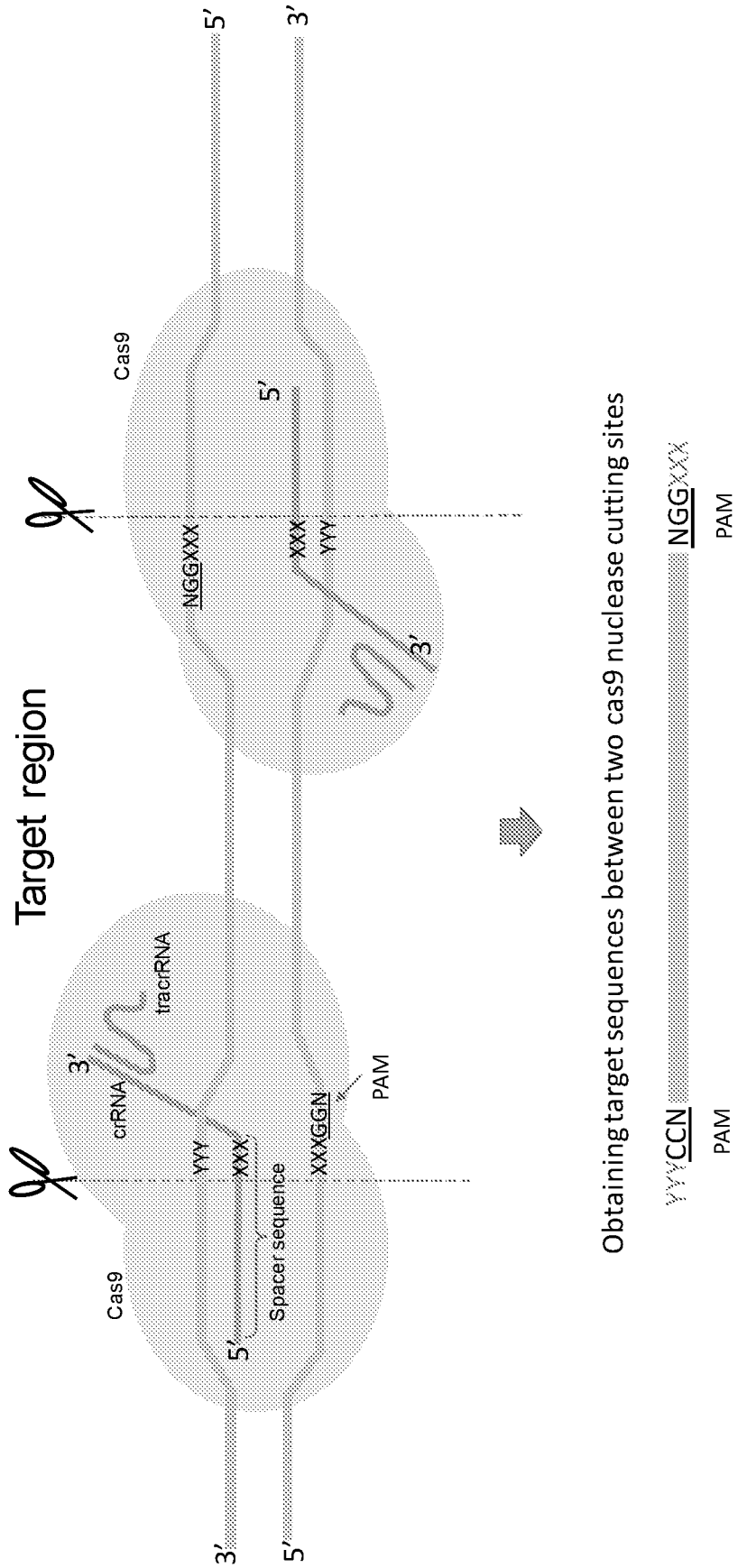

Target region
Cas9
crRNA
tracrRNA
Spacer sequence
PAM
XXXGGN
NGGXXX
Obtaining target sequences between two cas9 nuclease cutting sites
YYYCCN
NGGXXX
PAM
PAM

FIG. 93

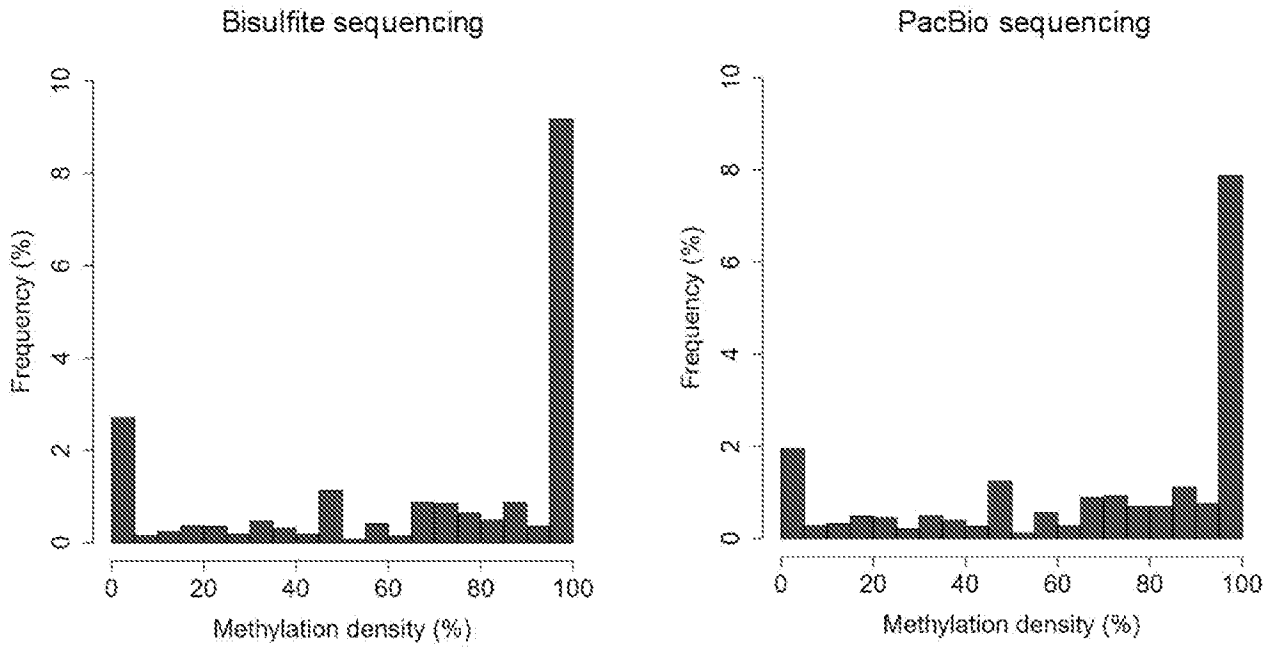

Bisulfite sequencing
PacBio sequencing
Frequency (%)
Methylation density (%)
0
2
4
6
8
10
0
20
40
60
80
100

FIG. 94

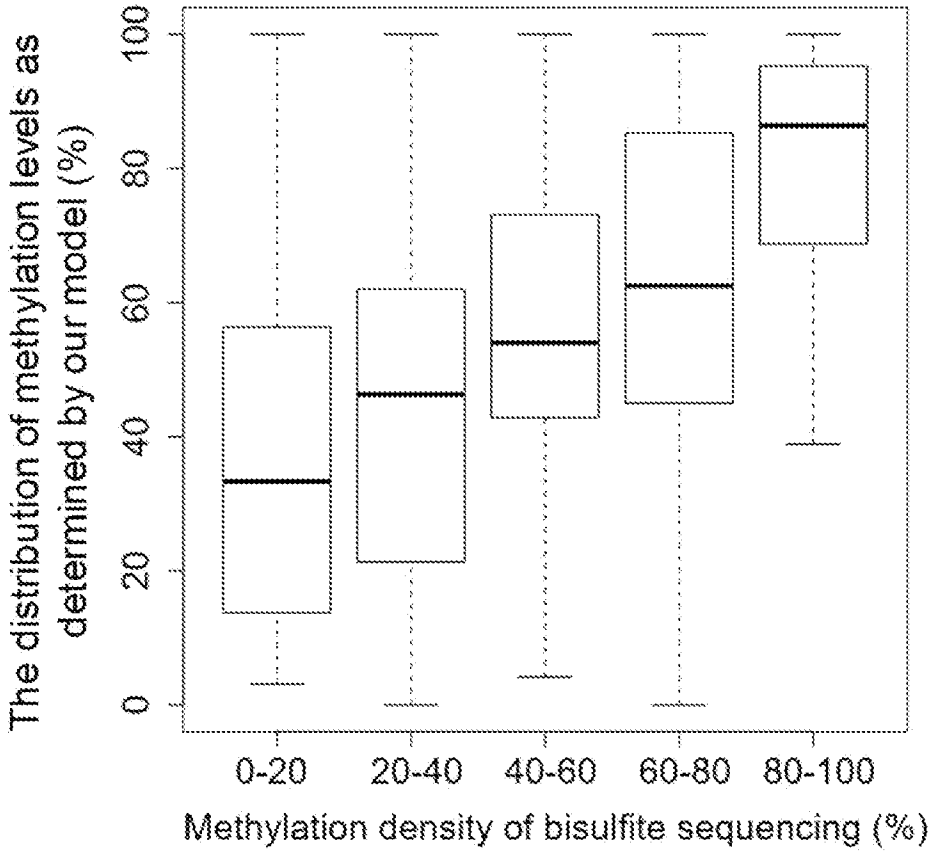
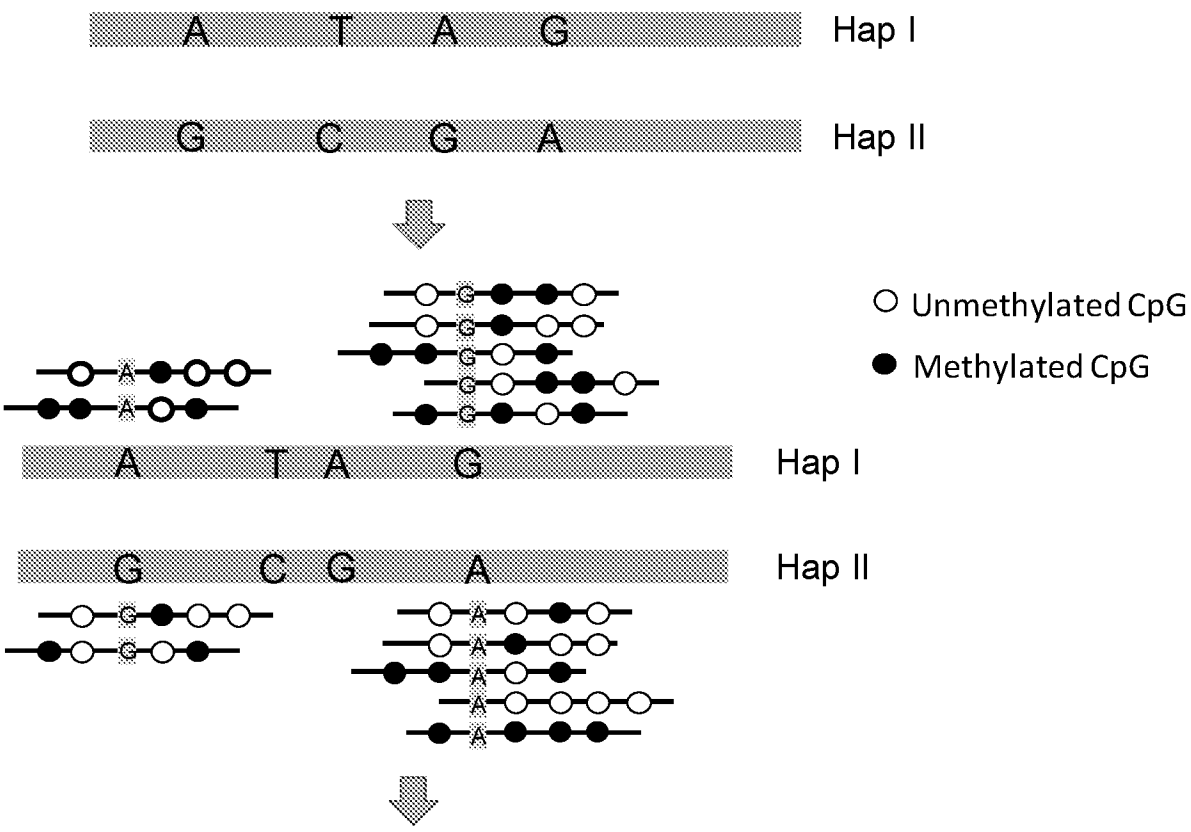

The distribution of methylation levels as determined by our model (%)
0
20
40
60
80
100
0-20
20-40
40-60
60-80
80-100
Methylation density of bisulfite sequencing (%)

FIG. 95



Relative haplotype-based methylation imbalance index

e.g. Methylation level of Hap I – Methylation level of Hap II

FIG. 104

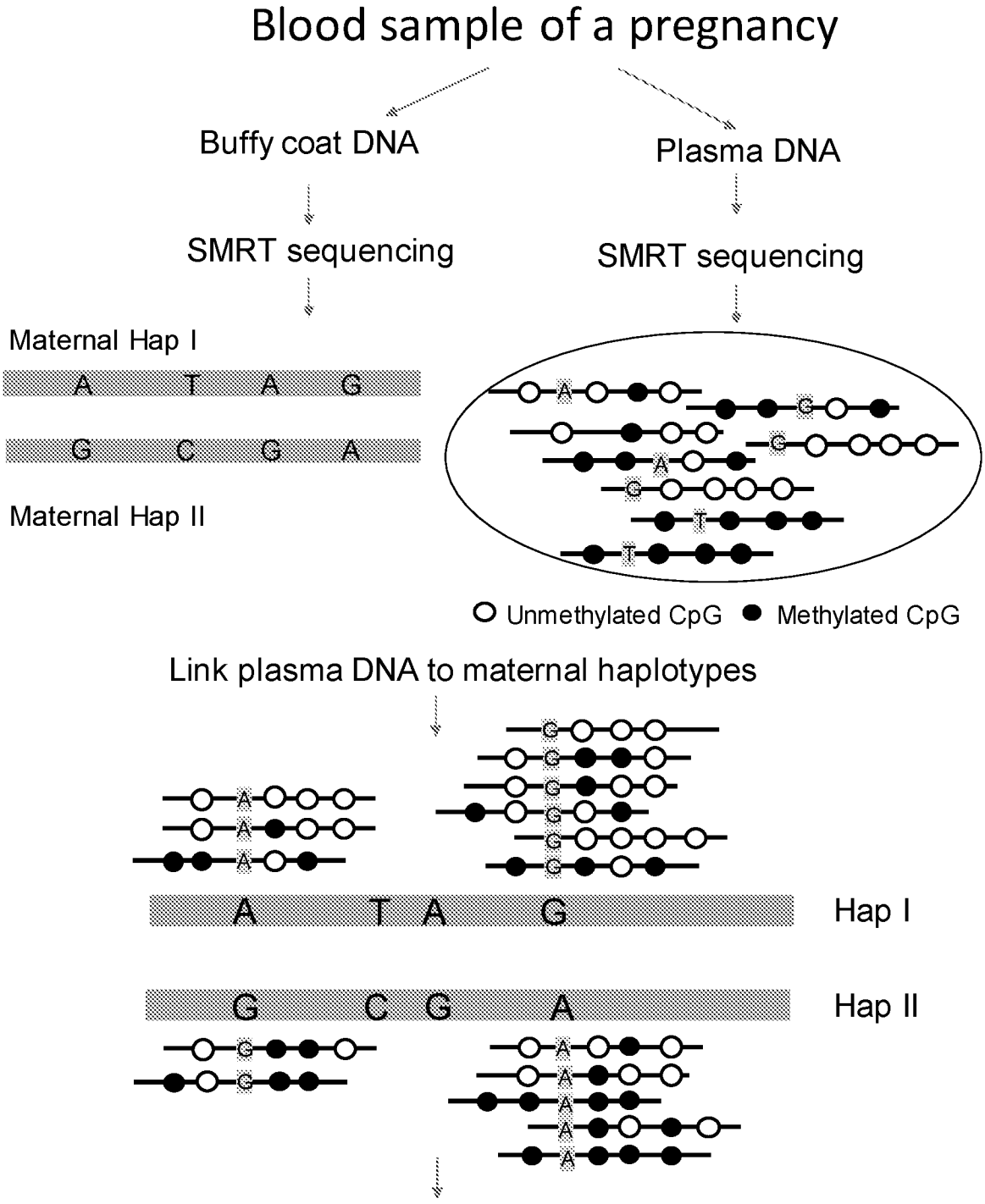

Blood sample of a pregnancy
Buffy coat DNA
Plasma DNA
SMRT sequencing
SMRT sequencing
Maternal Hap I
A T A G
G C G A
Maternal Hap II
Unmethylated CpG
Methylated CpG
Link plasma DNA to maternal haplotypes
A T A G
Hap I
G C G A
Hap II
Methylation level of Hap I < Methylation level of Hap II
The fetus has inherited Hap I

FIG. 108

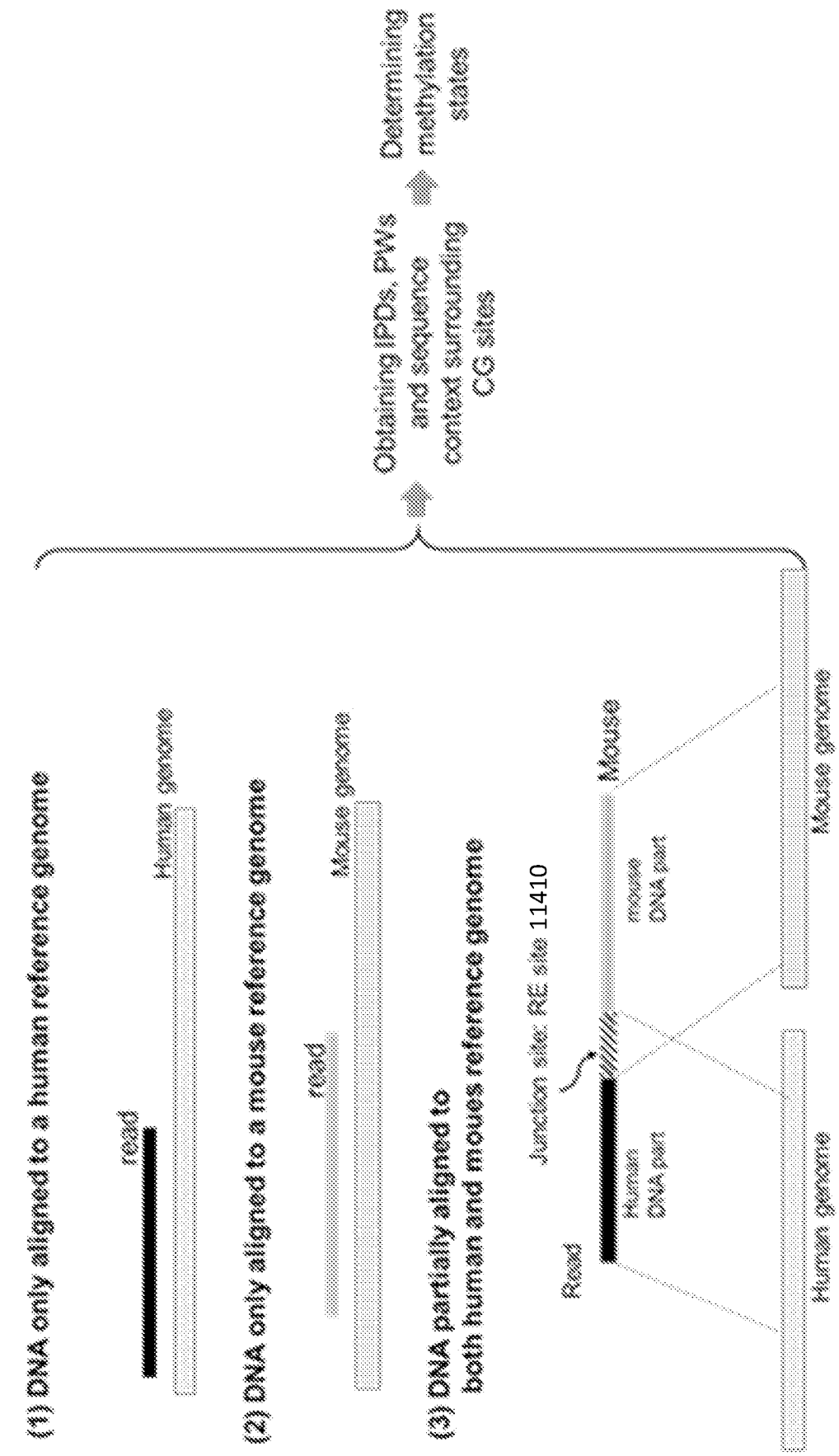

(1) DNA only aligned to a human reference genome
read
Human genome
(2) DNA only aligned to a mouse reference genome
read
Mouse genome
(3) DNA partially aligned to both human and moues reference genome
Junction site: RE site 11410
Read
Human DNA part
Mouse DNA part
Mouse
Mouse genome
Human genome
Obtaining IPDs, PWs and sequence context surrounding CG sites
Determining methylation states

FIG. 114

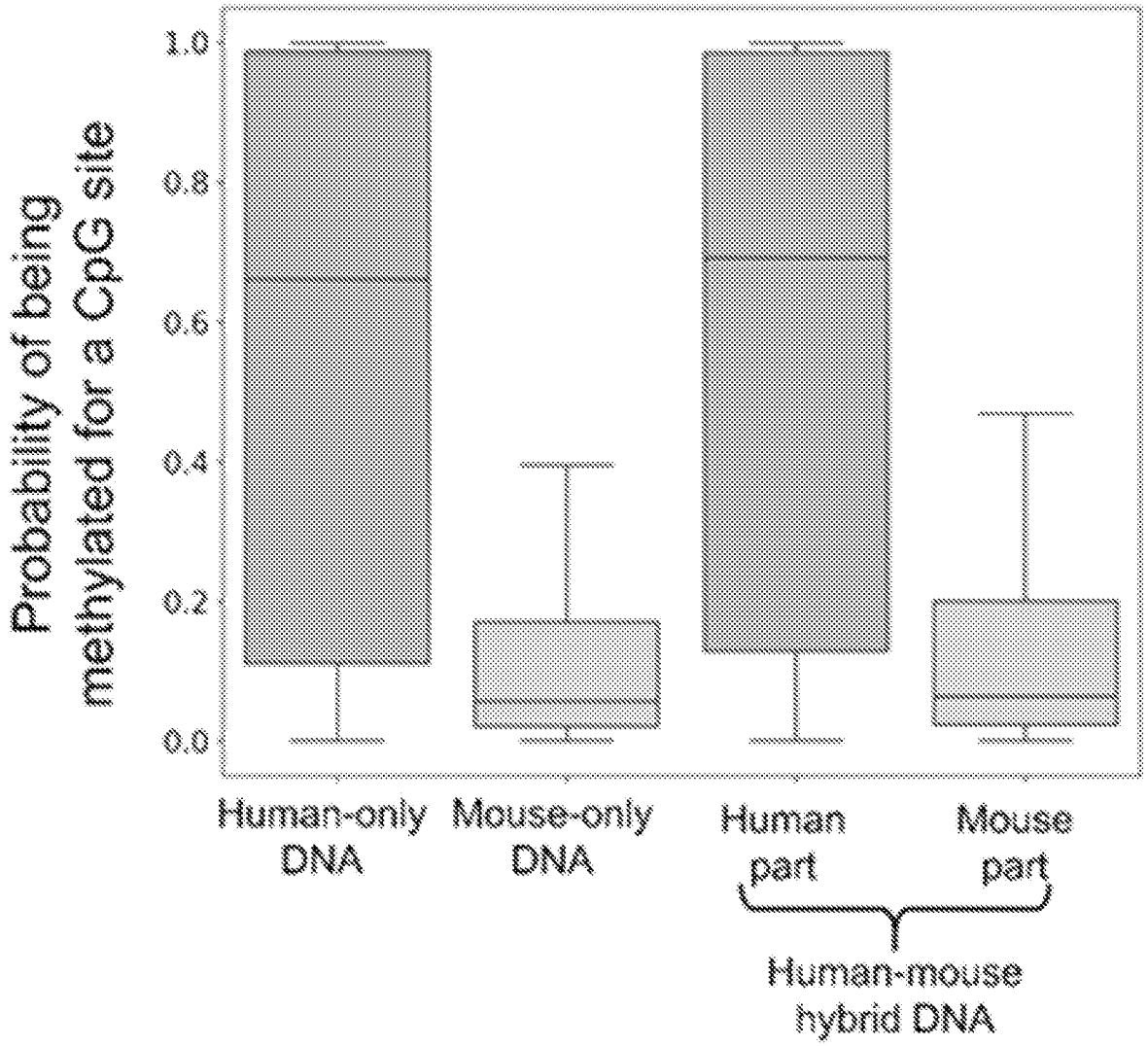

Probability of being methylated for a CpG site
1.0
0.8
0.6
0.4
0.2
0.0
Human-only DNA
Mouse-only DNA
Human part
Mouse part
Human-mouse hybrid DNA

FIG. 115

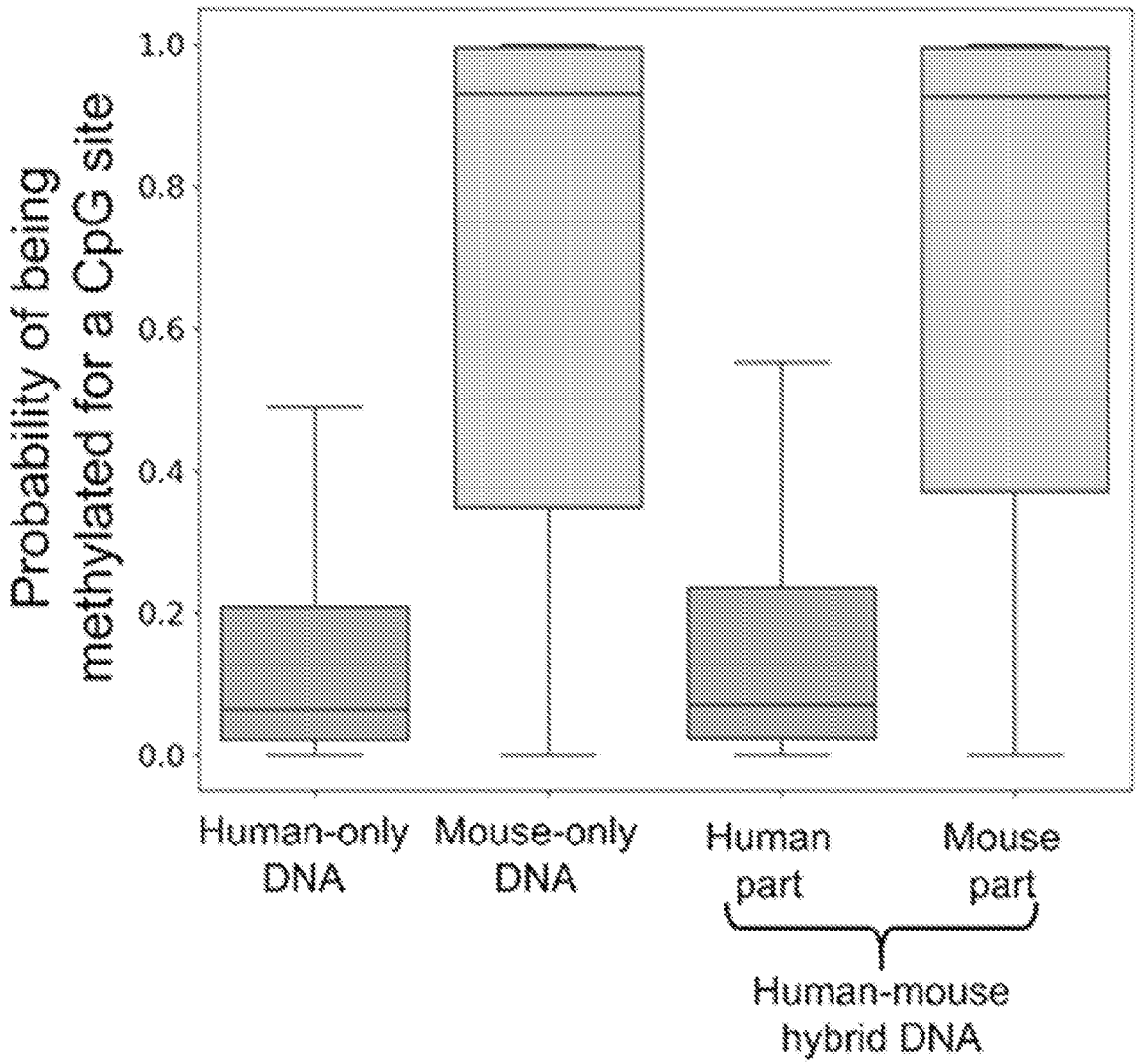

Probability of being methylated for a CpG site
1.0
0.8
0.6
0.4
0.2
0.0
Human-only DNA
Mouse-only DNA
Human part
Mouse part
Human-mouse hybrid DNA

FIG. 117

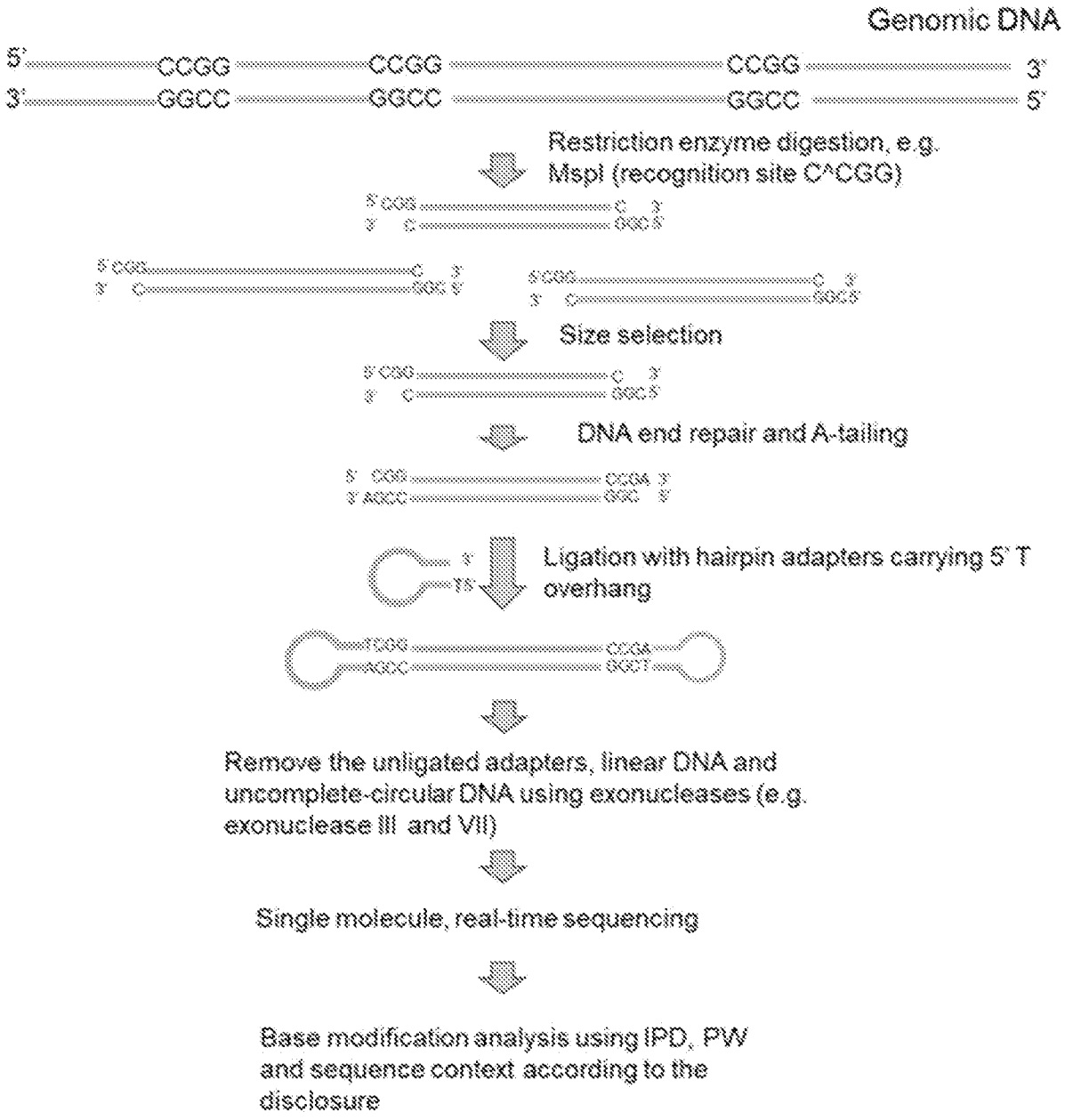

Genomic DNA
5'
CCGG
CCGG
CCGG
3'
3'
GGCC
GGCC
GGCC
5'
Restriction enzyme digestion, e.g. MspI (recognition site C^CGG)
Size selection
DNA end repair and A-tailing
Ligation with hairpin adapters carrying 5' T overhang
Remove the unligated adapters, linear DNA and uncomplete-circular DNA using exonucleases (e.g. exonuclease III and VII)
Single molecule, real-time sequencing
Base modification analysis using IPD, PW and sequence context according to the disclosure

FIG. 126

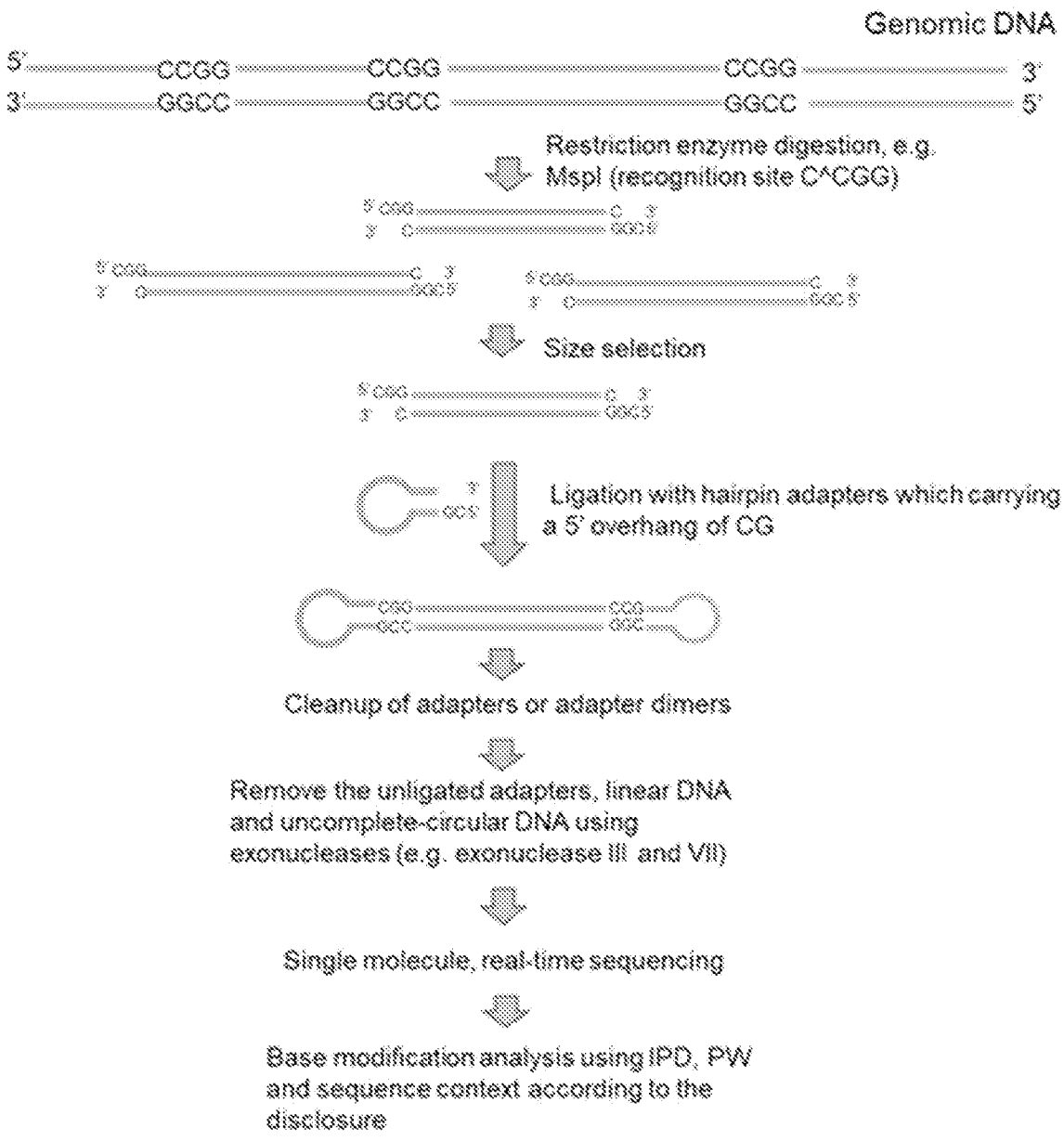

Genomic DNA
5' CCGG CCGG CCGG 3'
3' GGCC GGCC GGCC 5'
Restriction enzyme digestion, e.g. MspI (recognition site C^CGG)
Size selection
Ligation with hairpin adapters which carrying a 5' overhang of CG
Cleanup of adapters or adapter dimers
Remove the unligated adapters, linear DNA and uncomplete-circular DNA using exonucleases (e.g. exonuclease III and VII)
Single molecule, real-time sequencing
Base modification analysis using IPD, PW and sequence context according to the disclosure

FIG. 130

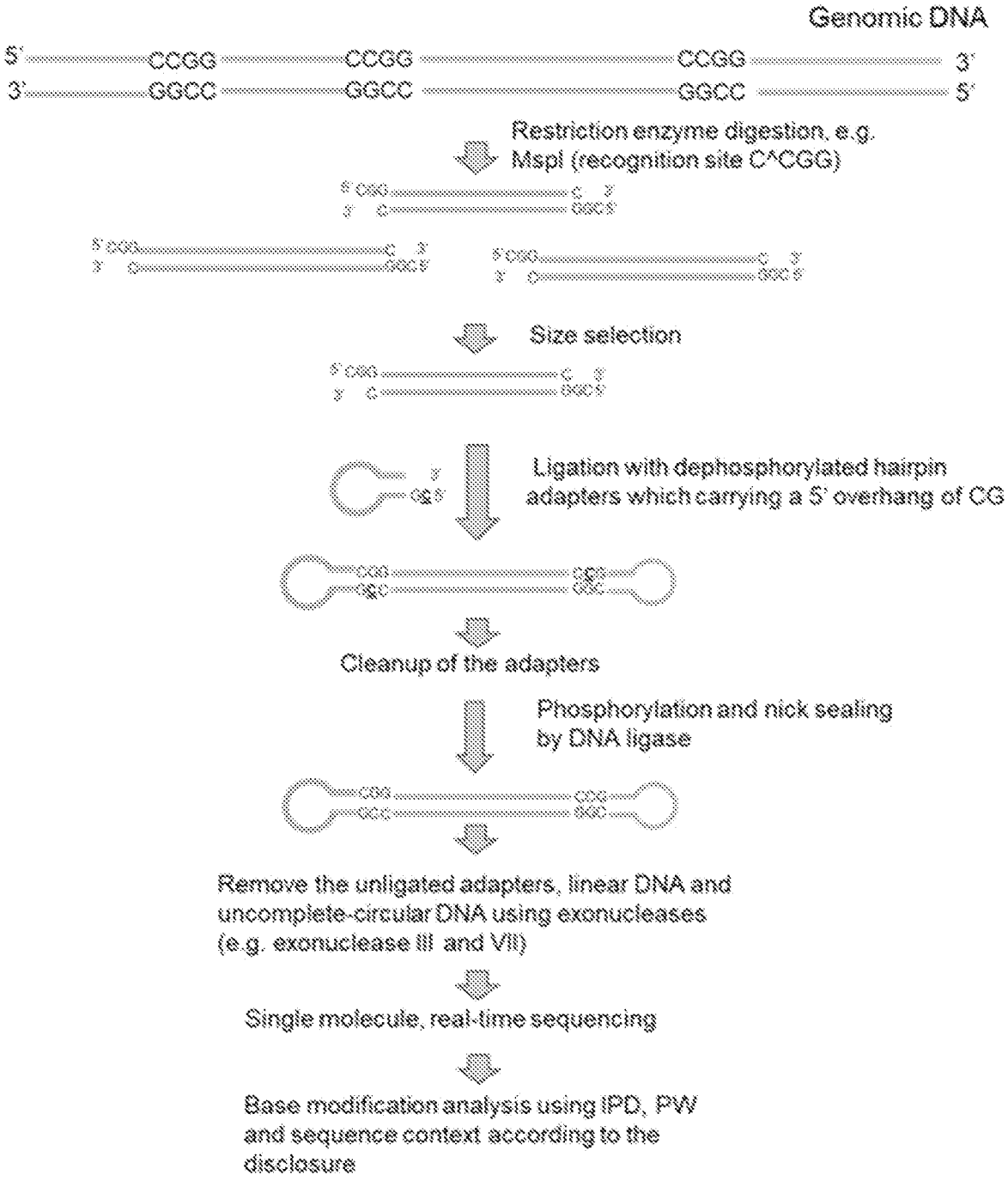

Genomic DNA
5' CCGG CCGG CCGG 3'
3' GGCC GGCC GGCC 5'
Restriction enzyme digestion, e.g.
MspI (recognition site C^CGG)
Size selection
Ligation with dephosphorylated hairpin
adapters which carrying a 5' overhang of CG
Cleanup of the adapters
Phosphorylation and nick sealing
by DNA ligase
Remove the unligated adapters, linear DNA and
uncomplete-circular DNA using exonucleases
(e.g. exonuclease III and VII)
Single molecule, real-time sequencing
Base modification analysis using IPD, PW
and sequence context according to the
disclosure

FIG. 131

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. 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.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 17/379,544	Filing Date 07/19/2021	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $100 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $480 =	
☐ 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).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	04/27/2022	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 23	Minus	** 30	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	*** 4	= 0		x $480 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/MARSHA R RICHARDS/
*** 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 number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

SCORE Placeholder Sheet for IFW Content

Application Number: 17379544 Document Date: 04/27/2022

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 DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

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.
17/379,544	07/19/2021	Yuk-Ming Dennis Lo	080015-028420US-1249377	2092

20350 7590 02/04/2022

Kilpatrick Townsend & Stockton LLP - West Coast
Mailstop: IP Docketing - 22
1100 Peachtree Street
Suite 2800
Atlanta, GA 30309

EXAMINER	
GODDARD, LAURA B	

ART UNIT	PAPER NUMBER
1642	

NOTIFICATION DATE	DELIVERY MODE
02/04/2022	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):

KTSDocketing2@kilpatrick.foundationip.com
ipefiling@kilpatricktownsend.com

Office Action Summary	Application No. 17/379,544	Applicant(s) Lo et al.	
	Examiner LAURA B GODDARD	Art Unit 1642	AIA (FITF) 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 3 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 January 19, 2022.

☐ 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) 1-30 is/are pending in the application.

5a) Of the above claim(s) 14-17 and 28-30 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 1-13 and 18-27 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ 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 July 19, 2021 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: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA.

1. The Election filed January 19, 2022 in response to the Office Action of November 23, 2021 is acknowledged. Applicant elected with traverse Group III, claims 18-27.

2. Applicants traverse the restriction between Groups III and I. The arguments are persuasive and **Group I, claims 1-13, are rejoined for examination**.

3. Claims 1-30 are pending. Claims 14-17 and 28-30 have been withdrawn from further consideration by the examiner under 35 CFR 1.142(b) as being drawn to non-elected inventions. Claims 1-13 and 18-27 are currently under prosecution.

Drawings/ Specification

Objection to Color Drawings/Specification

4. MPEP 608.02, part VIII states:

> *Color drawings and color photographs are not accepted in utility applications filed under 35 U.S.C. 111 unless a petition filed under 37 CFR 1.84(a)(2) or (b)(2) is granted. Color drawings and color photographs are not permitted in international applications (see PCT Rule 11.13).*
>
> *Unless a petition is filed and granted, color drawings or color photographs will not be accepted in a utility patent application filed under 35 U.S.C. 111. The examiner must object to the color drawings or color photographs as being*

improper and require applicant either to cancel the drawings or to provide substitute black and white drawings.

Since the color petition of July 19, 2021 was not granted in the decision of November 23, 2021, the color drawings or color photographs are being objected to as being improper. Unless Applicant submits a renewed petition that is granted, Applicant is required to either cancel the drawings or to provide substitute black and white drawings. The specification is also objected to for reciting the requisite color petition language as the first paragraph of the brief description of the drawings in which the color petition of July 19, 2021 was not granted. Unless Applicant submits a renewed petition that is granted, Applicant is required to amend the specification to remove the noted language, and also, if necessary, amend the section under brief description of the drawings to remove any language referencing color depiction in the drawings.

Double Patenting

The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed.

Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on nonstatutory double patenting provided the reference application or patent either is shown to be commonly owned with the examined application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. See MPEP § 717.02 for applications subject to examination under the first inventor to file provisions of the AIA as explained in MPEP § 2159. See MPEP § 2146 *et seq.* for applications not subject to examination under the first inventor to file provisions of the AIA. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO Internet website contains terminal disclaimer forms which may be used. Please visit www.uspto.gov/patent/patents-forms. The filing date of the application in which the form is filed determines what form (e.g., PTO/SB/25, PTO/SB/26, PTO/AIA/25, or PTO/AIA/26) should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved immediately upon submission. For more information about eTerminal Disclaimers, refer to www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

5. Claims 1-13 and 18-27 are rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1-19 of **U.S. Patent No. 11,091,794**.

Although the claims at issue are not identical, they are not patentably distinct from each other because the US Patent claims and instant claims are drawn to a method for detecting a modification of a nucleotide in a nucleic acid molecule, the method comprising:

(a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in a sample nucleic acid molecule and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(b) creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(c) inputting the input data structure into a model, the model trained by:

receiving a first plurality of first data structures, each first data structure of the first plurality of data structures corresponding to a respective window of nucleotides

sequenced in a respective nucleic acid molecule of a plurality of first nucleic acid molecules, wherein each of the first nucleic acid molecules is sequenced by measuring pulses in the optical signal corresponding to the nucleotides, wherein the modification has a known first state in a nucleotide at a target position in each window of each first nucleic acid molecule, each first data structure comprising values for the same properties as the input data structure,

storing a plurality of first training samples, each including one of the first plurality of first data structures and a first label indicating the first state of the nucleotide at the target position, and

optimizing, using the plurality of first training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels of the first labels when the first plurality of first data structures is input to the model, wherein an output of the model specifies whether the nucleotide at the target position in the respective window has the modification,

(d) determining, using the model, whether the modification is present in a nucleotide at the target position within the window in the input data structure;

wherein the nucleic acid molecules are obtained from a biological sample of a subject, wherein modification detected is a maternal or fetal methylation profile; wherein the sample nucleic acid molecule is a circular DNA molecule formed by cutting a dsDNA molecules using a Cas9 complex to form a cut dsDNA molecules and ligating a hairpin adaptor onto an end of the cut dsDNA molecules; wherein the nucleotides within the window are determined without using circular consensus sequences and without

alignment to a reference genome; wherein the optical signal is a fluorescence signal from a dye-labeled nucleotide.

6. **Conclusion:** No claim is allowed. The closes prior art made of record but not relied upon are US Patent 9,175,348; US Patent 9,175,338; and US Patent 10,590,484. All three US Patents teach methods for single molecule real time (SMRT) sequencing to identify methylation in sequences. The US Patents teach measuring pulse width of nucleotides, interpulse duration (IPD), and aligning sequences and signals to identify differences in IPD and pulse width as indicators of genetic alteration such as methylation (Figure 2 in US Patent 9,175,348 and US Patent 9,175,338). US Patent 10,590,484 additionally teaches conducting these methods on samples from fetal or maternal origin. None of the US Patents demonstrate measuring a pulse width value of a specifically identified nucleotide at an identified position to determine the methylation status, and do not teach or suggest combining all data measurements of nucleotide identities, nucleotide positions in the sample nucleic acid, width of the pulse corresponding to the nucleotides, and IPD, subsequently creating an input data structure comprising a window of the nucleotides in a sample that lists the above measurements, and inputting the measurements into a model configured to utilize the measurements to output a methylation status for each nucleotide as instantly claimed.

7. Any inquiry concerning this communication or earlier communications from the examiner should be directed to LAURA B GODDARD whose telephone number is (571)272-8788. The examiner can normally be reached Mon-Fri, 7am-3:30pm.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Misook Yu can be reached on 571-272-0839. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of published or unpublished applications may be obtained from Patent Center. Unpublished application information in Patent Center is available to registered users. To file and manage patent submissions in Patent Center, visit: https://patentcenter.uspto.gov. Visit https://www.uspto.gov/patents/apply/patent-center for more information about Patent Center and https://www.uspto.gov/patents/docx for information about filing in DOCX format. For additional questions, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Laura B Goddard/
Primary Examiner, Art Unit 1642

Notice of References Cited	Application/Control No. 17/379,544	Applicant(s)/Patent Under Reexamination Lo et al.	
	Examiner LAURA B GODDARD	Art Unit 1642	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-9175348-B2	11-2015	Korlach; Jonas	C12Q1/6827	1/1
*	B	US-9175338-B2	11-2015	Flusberg; Benjamin	C12Q1/6869	1/1
*	C	US-10590484-B2	03-2020	Korlach; Jonas	C12Q1/6874	1/1
	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.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	17/379,544	Lo et al.
	Examiner	Art Unit
	LAURA B GODDARD	1642

CPC - Searched*		
Symbol	Date	Examiner

CPC Combination Sets - Searched*		
Symbol	Date	Examiner
(C12Q1/6816 or C12N9/22 or G16B20/00 or G16B30/00 orC12Q1/6851)	01/31/2022	LBG

US Classification - Searched*			
Class	Subclass	Date	Examiner

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
WEST	01/31/2022	LBG
STN	01/31/2022	LBG
Review prosecution 16/995,607	01/31/2022	LBG

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

/Laura B Goddard/ Primary Examiner, Art Unit 1642	

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	17/379,544
Filing Date	July 19, 2021
First Named Inventor	Yuk-Ming Dennis Lo
Art Unit	1642
Examiner Name	Laura B. Goddard
Attorney Docket Number	080015-1249377

U.S. PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code [1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1	8467976		Jun 18, 2013	Lo et al.	
	2	9116118		Aug 25, 2015	Turner et al.	
	3	9175338		Nov 3, 2015	Flusberg et al.	
	4	9175341		Nov 3, 2015	Flusberg et al.	
	5	10438691		Oct 8, 2019	Kim et al.	
	6	10465232		Nov 5, 2019	Wu et al.	
	7	10468121		Nov 5, 2019	Kermani et al.	
	8	10590484		Mar 17, 2020	Korlach et al.	
	9	11091794		Aug 17, 2021	Lo et al.	

U.S. PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code [1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	10	20110183320		Jul 28, 2011	Flusberg et al.	
	11	20130230909		Sep 5, 2013	Pan et al.	
	12	20140004511		Jan 2, 2014	Korlach et al.	
	13	20160017419		Jan 21, 2016	Chiu et al.	
	14	20170037464		Feb 9, 2017	Turner et al.	
	15	20170159132		Jun 8, 2017	Okino et al.	
	16	20170233802		Aug 17, 2017	Flusberg et al.	
	17	20180201993		Jul 19, 2018	Turner et al.	
	18	20190024162		Jan 24, 2019	Korlach et al.	

FOREIGN PATENT DOCUMENTS

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T[5]
	19	2009013492	WO		Jan 29, 2009	The Chinese University of Hong Kong		
	20	2010027484	WO		Mar 11, 2010	Pacific Biosciences of California Inc.		
	21	2010068289	WO		Jun 17, 2010	Pacific Biosciences of California Inc.		
	22	2013170429	WO		Nov 21, 2013	BGI Genomics Co. Ltd.	English Abstract Attached	
	23	2013188846	WO		Dec 19, 2013	Harry Stylli		
	24	104053789	CN		Sep 17, 2014	BGI Genomics Co. Ltd.	English Abstract Attached	
	25	2014153757	WO		Oct 2, 2014	BGI Shenzhen	English Abstract Attached	
	26	104053789	CN		Feb 10, 2016	BGI Genomics Co. Ltd.	English Abstract Attached	
	27	2012380221	AU		Sep 29, 2016	BGI Genomics Co. Ltd.		
	28	2017012544	WO		Jan 26, 2017	The Chinese University of Hong Kong		
	29	101896167	KR		Sep 7, 2018	The Chinese University of Hong Kong	English Abstract of corresponding document: WO2009013492 (A1) Attached	

NON-PATENT LITERATURE DOCUMENTS

Examiner 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.	T[5]
	30	BEAULAURIER et al., Deciphering Bacterial Epigenomes Using Modern Sequencing Technologies, Nature Reviews Genetics, Volume 20 Issue 3, March 2019, pp. 157-172.	
	31	BLOW et al., The Epigenomic Landscape of Prokaryotes, PLOS Genetics, Volume 12, No. 2, February 12, 2016, pp. 1-28.	
	32	CLARK et al., Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing, Nucleic Acids Research, Volume 40, No. 4, Available online at: https://doi.org/10.1093/nar/gkr1146, 2012, 12 pages.	
	33	CLARK et al., Enhanced 5-Methylcytosine Detection in Single-Molecule, Real-Time Sequencing via Tet1 Oxidation, BMC Biology, Volume 11, No. 4, 2013, 10 pages.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

	34	EID et al., Real-Time DNA Sequencing from Single Polymerase Molecules, Science, Volume 323, No. 5910, January 2, 2009, pp. 133-138.	
	35	FENG et al., Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic, PLOS Computational Biology, Volume 9, No. 3, Available online at: https://doi.org/10.1371/journal.pcbi.1002935, March 2013, 10 pages.	
	36	FENG et al., qDNAmod: A Statistical Model-based Tool to Reveal Intercellular Heterogeneity of DNA Modification from SMRT Sequencing Data, Nucleic Acids Research, Volume 42, No. 22, November 17, 2014, pp. 13488-13499.	
	37	FLUSBERG et al., Direct Detection of DNA Methylation During Single-Molecule, Real-Time Sequencing, Nature Methods, Volume 7, No. 6, June 2010, pp. 461-465.	
	38	HE et al., TETI Promotes Leukemic Growth of AMLI-ETO+ AML Via 5-Hydroxymethylcytosine Marks and its Oncogenic Role and High Expression can be Antagonized Using Olaparib, an Inhibitor of its Binding Partner PARPI, Experimental Hematology, Volume 53, 2017, pp. S92-S93.	
	39	KILE et al., Correlation of Global and Gene-Specific DNA Methylation in Maternal-Infant Pairs, PLoS One, Volume 5, No. 10, October 29, 2010, 9 pages.	
	40	LIU et al., Bisulfite-free Direct Detection of 5-Methylcytosine and 5- Hydroxymethylcytosine at Base Resolution, Nature Biotechnology Letters, Volume 37, April 2019, pp. 424-429.	
	41	LIU et al., Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data, Nature Communications, Volume 10, No. 2449, June 4, 2019, 11 pages.	
	42	NI et al., Deepsignal: Detecting DNA Methylation State From Nanopore Sequencing Reads Using Deep-Learning, Bioinformatics, Volume 35, No. 22, 2019, pp. 4586-4595.	
	43	SCHADT et al., Modeling Kinetic Rate Variation in Third Generation DNA Sequencing Data to Detect Putative Modifications to DNA Bases, Genome Research, Volume 23, No. 1, January 2013, pp. 129-141.	
	44	SUGAI et al., Analysis of the DNA Methylation Level of Cancer-related Genes in Colorectal Cancer and the Surrounding Normal Mucosa, Clinical Epigenetics, Volume 9, No. 55, May 18, 2017, pp. 1-10.	
	45	SUZUKI et al., Agin: Measuring the Landscape of CpG Methylation of Individual Repetitive Elements, Bioinformatics, Volume 32, No.19, June 17, 2016, pp. 2911-2919.	
	46	SUZUKI, Observing Comprehensive DNA Methylomes via Single Molecule Real-Time Sequencing: Application to Diploid and Centromeric Methylation, Dissertation submitted to The Graduate School of Frontier Sciences, The University of Tokyo, December 13, 2017, 118 pages.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

EXAMINER SIGNATURE			
Examiner Signature	/Laura B Goddard/	Date Considered	01/31/2022

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document.
[4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David D. Hsu/	Date (YYYY-MM-DD)	2021-12-22
Name/Print	David D. Hsu	Registration Number	71,659

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /LBG/

WEST Search History

Hide Items | Restore | Clear | Close

DATE: Monday, January 31, 2022

Hide?	Set Name	Query	Hit Count
DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI; PLUR=YES; OP=OR			
☐	L17	L16 AND ((pulse interpulse peak) NEAR (width distance))	5
☐	L16	L2 AND "Chinese University of Hong Kong"	26
☐	L15	L14 AND (fetal fetus maternal pregnant mother)	497
☐	L14	(C12Q1/6816 or C12N9/22 or G16B20/00 or G16B30/00 or C12Q1/6851).cpc. AND (SMRT "single molecule real time sequencing")	1018
☐	L13	L12 AND width[clm]	1
☐	L12	L8 AND (width position)[CLM]	34
☐	L11	L8 AND (interpulse pulse duration)[CLM]	0
☐	L10	L8 AND (position AND (distance width) AND (interpulse pulse duration))[CLM]	0
☐	L9	L4 AND (fetal fetus maternal pregnant mother)[CLM]	20
☐	L8	L3 AND (fetal fetus maternal pregnant mother)[CLM]	296
☐	L7	L5 AND Cas9	31
☐	L6	L5 AND interpulse	13
☐	L5	L4 AND (methylation methylated)	73
☐	L4	L2 AND position AND width AND pulse	115
☐	L3	L2 AND @ay<=2019	2248
☐	L2	("single molecule real time sequencing" SMRT) AND (fetal fetus maternal pregnant mother)	2543
☐	L1	("single molecule real time sequencing" SMRT) AND (fetal fetus maternal mother)	2493

END OF SEARCH HISTORY

```
     (FILE 'HOME' ENTERED AT 11:20:16 ON 31 JAN 2022)
               D CLUSTER
               D CLUSTER .BIO
               D SET

     FILE 'CAPLUS, EMBASE, BIOSIS, MEDLINE, SCISEARCH' ENTERED AT 11:20:40 ON
     31 JAN 2022
L1          135 SEA SPE=ON  ABB=ON  PLU=ON  ("SINGLE MOLECULE REAL TIME
                SEQUENCING" OR SMRT) AND (FETAL OR FETUS OR MATERNAL OR
                PREGNANT)
L2          112 SEA SPE=ON  ABB=ON  PLU=ON  L1 AND PY<2020
L3            0 SEA SPE=ON  ABB=ON  PLU=ON  L2 AND (PULSE OR PEAK)(S)(WIDTH OR
                DISTANCE)
L4            0 SEA SPE=ON  ABB=ON  PLU=ON  L2 AND (PULSE OR PEAK) AND (WIDTH
                OR DISTANCE)
L5            0 SEA SPE=ON  ABB=ON  PLU=ON  L2 AND PULSE
```

I hereby certify that this correspondence is being filed via EFS-Web with the United States Patent and Trademark Office on January 19, 2022.

KILPATRICK TOWNSEND & STOCKTON LLP

By: /Heather A. Orosco/
Heather A. Orosco

PATENT
Attorney Docket No.: 080015-028420US-1249377
Client Ref. No.: 19/MED/901 US Con1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Confirmation No.: 2092
Yuk-Ming Dennis Lo et al.	Examiner: Laura B. Goddard
Application No.: 17/379,544	Technology Center/Art Unit: 1642
Filed: July 19, 2021	
For: DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS	**AMENDMENT / RESPONSE TO RESTRICTION REQUIREMENT**
Customer No.: 20350	

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

Commissioner:

In response to the Restriction Requirement mailed November 23, 2021, in the above-referenced application, please enter the following amendments and consider the following remarks:

Amendments to the Claims are reflected in the listing of claims, which begins on page 2 of this paper.

Remarks/Arguments begin on page 13 of this paper.

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings of claims in the application.

Listing of Claims:

1. (Withdrawn – Currently Amended) A method for determining whether a sample nucleic acid molecule is of fetal or maternal origin, wherein the sample nucleic acid molecule is cell-free and obtained from a biological sample of a female subject pregnant with a fetus, the method comprising:

(a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in the sample nucleic acid molecule and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(b) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:

receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,

apply optimized parameters of the model to the properties of the input data structure; and

output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation; [[and]]

(d) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, the methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(e) determining whether the sample nucleic acid molecule is of fetal or maternal origin using the methylation statuses of the one or more nucleotides.

2. (Withdrawn) The method of claim 1, wherein determining whether the sample nucleic acid molecule is of fetal or maternal origin comprises:

determining a methylation level of the sample nucleic acid molecule using the methylation statuses of the one or more nucleotides; and

comparing the methylation level of the sample nucleic acid molecule to a reference value.

3. (Withdrawn) The method of claim 2, wherein the reference value is determined from a methylation level of one or more maternal nucleic acid molecules.

4. (Withdrawn) The method of claim 2, wherein:

comparing the methylation level of the sample nucleic acid molecule to the reference value comprises determining the methylation level of the sample nucleic acid molecule is lower than the reference value, and

determining whether the sample nucleic acid molecule is of fetal or maternal

origin comprises determining the sample nucleic acid molecule is of fetal origin using the comparison.

5. (Withdrawn) The method of claim 1, further comprising:

identifying the sample nucleic acid molecule as aligning to a predetermined genomic region.

6. (Withdrawn) The method of claim 1, wherein:

the sample nucleic acid molecule is one sample nucleic acid molecule of a plurality of sample nucleic acid molecules,

the method further comprising:

performing steps (a)-(e) for each of the plurality of sample nucleic acid molecules to determine whether each of the plurality of sample nucleic acid molecules is of fetal or maternal origin using the methylation statuses, and

determining a fetal fraction using the determination of the fetal or maternal origin of the plurality of sample nucleic acid molecules.

7. (Withdrawn) The method of claim 1, wherein the model includes a machine learning model, a principal component analysis, a convolutional neural network, or a logistic regression.

8. (Withdrawn) The method of claim 1, wherein the nucleotides within the window are determined using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

9. (Withdrawn) The method of claim 1, wherein nucleotides within the window are determined without using a circular consensus sequence and without alignment of the sequenced nucleotides to a reference genome.

10. (Withdrawn) The method of claim 1, wherein the optical signal is a fluorescence signal from a dye-labeled nucleotide.

11. (Withdrawn) The method of claim 1, wherein each window associated

with each input data structure comprises at least 4 consecutive nucleotides on a first strand of each first nucleic acid molecule.

12. (Withdrawn) The method of claim 1, wherein the methylation is 5mC or 6mA.

13. (Withdrawn) The method of claim 1, wherein:

determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, and

the sample nucleic acid molecule is a first nucleic acid molecule;

the method further comprising:

providing a second nucleic acid molecule, the second nucleic acid molecule being identical to the first nucleic acid molecule and having a methylated nucleotide at a position corresponding to the target position,

performing bisulfite sequencing on the second nucleic acid molecule to obtain bisulfite sequencing results, and

verifying the methylation present in the nucleotide at the target position using bisulfite sequencing results.

14. (Withdrawn) A method of determining whether a copy number aberration is present in a region of a fetal genome, the method comprising:

for each sample nucleic acid molecule of a plurality of sample nucleic acid molecules, wherein the plurality of sample nucleic acid molecules is cell-free and obtained from a biological sample of a female subject pregnant with a fetus:

(a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in the sample nucleic acid molecule and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid

molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(b) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:

receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,

apply optimized parameters of the model to the properties of the input data structure; and

output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation,

(d) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, a methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(e) identifying the sample nucleic acid molecule as aligning to the region; determining a methylation level of the region using the plurality of methylation

statuses of the one or more nucleotides for each sample nucleic acid molecule of the plurality of sample nucleic acid molecules; and

determining whether the copy number aberration is present at the region of the fetal genome using the methylation level.

15. (Withdrawn) The method of claim 14, further comprising comparing the methylation level of the region to a reference level, wherein determining whether the copy number aberration is present at the region comprises using the comparison.

16. (Withdrawn) The method of claim 15, wherein the reference level is determined using a region without the same type of copy number aberration.

17. (Withdrawn) The method of claim 14, wherein the region is a chromosome,

the method further comprising:

determining the copy number aberration is present, and

determining the fetus has a chromosomal aneuploidy.

18. (Currently Amended) A method for determining whether a sample nucleic acid molecule is of fetal or maternal origin, wherein the sample nucleic acid molecule is cell-free and obtained from a biological sample of a female subject pregnant with a fetus, the method comprising:

(a) sequencing the sample nucleic acid molecule by measuring pulses in an optical signal corresponding to nucleotides in the sample nucleic acid molecule;

(b) receiving data acquired by sequencing the sample nucleic acid molecule, and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse

corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(d) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:

receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,

apply optimized parameters of the model to the properties of the input data structure; and

output a methylation status specifying whether the nucleotide at the target position in the respective window has a methylation; [[and]]

(e) for each of the one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, the methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

([[e]]f) determining whether the sample nucleic acid molecule is of fetal or maternal origin using the methylation statuses of the one or more nucleotides.

19. (Original) The method of claim 18, further comprising:

cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and

ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule

to form a circular DNA molecule, wherein:

the circular DNA molecule is the sample nucleic acid molecule,

the window of the nucleotides corresponding to the input data structure comprises nucleotides on a first strand of the sample nucleic acid molecule and nucleotides on a second strand of the sample nucleic acid molecule, and

the input data structure further comprises for each nucleotide within the window a value of a strand property, the strand property indicating the nucleotide being present on either the first strand or the second strand.

20. (Original) The method of claim 18, further comprising enriching or filtering each nucleotide within the window.

21. (Original) The method of claim 20, comprising enriching each nucleotide within the window by cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule.

22. (Original) The method of claim 20, comprising filtering each nucleotide within the window selecting double-stranded DNA molecules having a size within a size range.

23. (Original) The method of claim 18, wherein sequencing the sample nucleic acid molecule comprises:

providing a polymerase that is capable of processing a first nucleic acid molecule, wherein the first nucleic acid molecule comprises the methylation in at least one nucleotide;

incorporating dye-labeled nucleotides into the first nucleic acid molecule to form the sample nucleic acid molecule; and

measuring pulses in the optical signal corresponding to the dye-labeled nucleotides,

wherein determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, thereby determining the methylation is present in the at least one nucleotide in the first nucleic acid molecule.

24. (Original) The method of claim 23, wherein the polymerase is immobilized.

25. (Original) The method of claim 23, further comprising ligating Watson and Crick strands of the first nucleic acid molecule to form a circular template molecule comprising a double-stranded segment having each end capped with a hairpin adaptor, wherein the circular template molecule is the first nucleic acid molecule.

26. (Original) The method of claim 18, further comprising:

providing a double-stranded nucleic acid molecule comprising the methylation in at least one nucleotide; and

ligating Watson and Crick strands of the double-stranded nucleic acid molecule to form a circular template molecule comprising a double-stranded segment having each end capped with a hairpin adaptor, wherein the circular template molecule is the sample nucleic acid molecule; wherein:

the data comprises sequences of the Watson and Crick strands of the double-stranded nucleic acid molecule contained in the circular template molecule, and

determining whether the methylation is present in the nucleotide at the target position comprises determining that the methylation is present in the nucleotide in the sample nucleic acid molecule at the target position, thereby determining the methylation is present in the at least one nucleotide in the double-stranded nucleic acid molecule.

27. (Currently Amended) The method of claim 18, further comprising aligning the sequenced nucleotides to a reference genome to provide the positions of the nucleotides within the sample nucleic acid molecule[[s]].

28. (Withdrawn) A method of determining whether a copy number aberration is present in a region of a fetal genome, the method comprising:

for each sample nucleic acid molecule of a plurality of sample nucleic acid molecules, wherein the plurality of sample nucleic acid molecules is cell-free and obtained from a biological sample of a female subject pregnant with a fetus:

(a) sequencing the sample nucleic acid molecule by measuring pulses in an optical signal corresponding to nucleotides in the sample nucleic acid molecule;

(b) receiving data acquired by sequencing the sample nucleic acid molecule, and obtaining, from the data, values for the following properties:

for each nucleotide:

an identity of the nucleotide,

a position of the nucleotide within the sample nucleic acid molecule,

a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(c) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(d) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, inputting the input data structure into a model, the model configured to:

receive the input data structure, generated using the data acquired by measuring the pulses in the optical signal corresponding to the nucleotides,

apply optimized parameters of the model to the properties of the input data structure; and

output a methylation status specifying whether the nucleotide at

the target position in the respective window has a methylation,

(e) for each of one or more nucleotides of the nucleotides sequenced in the sample nucleic acid molecule, determining, using the model, a methylation status for whether the methylation is present in the nucleotide when the nucleotide is at the target position within the window in the input data structure; and

(f) identifying the sample nucleic acid molecule as aligning to the region;

determining a methylation level of the region using the plurality of methylation statuses of the one or more nucleotides for each sample nucleic acid molecule of the plurality of sample nucleic acid molecules; and

determining whether the copy number aberration is present at the region of the fetal genome using the methylation level.

29. (Withdrawn) The method of claim 28, further comprising:

enriching the biological sample for nucleic acid molecules aligning to the region.

30. (Withdrawn) The method of claim 29, wherein enriching comprises cutting a double-stranded DNA molecule using a Cas9 complex to form a cut double-stranded DNA molecule, and ligating a hairpin adaptor onto an end of the cut double-stranded DNA molecule.

REMARKS/ARGUMENTS

Upon entry of the present amendment, claims 1-30 will be pending in this application. Claims 1, 18, and 27 have been amended to clarify antecedent basis. Claim 18 has also been amended to add a step of determining methylation status. Support can at least be found in paragraph [0517] and claim 1 of the Application as filed. No new matter is added.

The Examiner has restricted the claims into the following Groups:

Group I: Claims 1-13, drawn to a method for determining whether a sample nucleic acid molecule is of fetal or maternal origin;

Group II: Claims 14-17, drawn to a method of determining whether a copy number aberration (CNA) is present in a region of a fetal genome;

Group III: Claims 18-27, drawn to a method for determining whether a sample nucleic acid molecule is of fetal or maternal origin, with sequencing; and

Group IV: Claims 28-30, drawn to a method of determining whether a copy number aberration (CNA) is present in a region of a fetal genome, with sequencing.

In response to the restriction requirement, but without conceding the propriety thereof, Applicant elects Group III, encompassed by claims 18-27, with traverse, and without prejudice to the presentation of the claims of Groups I, II, and IV in later applications.

The election of Group III over Group I requirement is respectfully traversed. Related inventions are distinct if: (1) the inventions as claimed are either not capable of use together or can have a materially different design, mode of operation, function, or effect; (2) the inventions do not overlap in scope, i.e., are mutually exclusive; and (3) the inventions as claimed are not obvious variants. See MPEP § 806.05(j). Group III and Group I are both capable of use together and are not mutually exclusive. Claim 18 of Group III includes all steps of claim 1 of Group I. As a result, performing claim 18 involves performing claim 1. Claim 18 and claim 1 therefore are capable of use together and the claims are not mutually exclusive. Hence, the groups are not distinct.

Additionally, Group I would not involve a serious search and/or examination burden over Group III. Claim 18 of Group III includes all steps of claim 1 of Group I with the addition of a

sequencing step. Hence, there would be little or no additional search and/or examination for claim 1 after performing search and/or examination on claim 18.

For at least these reasons, an election of Group III over Group I should not be required.

CONCLUSION

In view of the foregoing, Applicant believes all claims now pending in this application are in condition for allowance. The issuance of a formal Notice of Allowance at an early date is respectfully requested.

Except for the issue fees payable under 37 C.F.R. § 1.18, the Director is authorized to charge any additional fees during pendency of this application, including any required extension of time fees, or credit any overpayment to Deposit Account Number 20-1430. This paragraph is intended to be a constructive petition for extension of time in accordance with 37 C.F.R. § 1.136(a)(3).

If the Examiner believes a telephone conference would expedite prosecution of this application, please contact the undersigned at (303) 571-4000 or dhsu@kilpatricktownsend.com.

Respectfully submitted,

/David D. Hsu/
David D. Hsu
Registration No. 71659

KILPATRICK TOWNSEND & STOCKTON LLP

DDH:hao

75444189V.1

Electronic Acknowledgement Receipt

EFS ID:	44783685
Application Number:	17379544
International Application Number:	
Confirmation Number:	2092
Title of Invention:	DETERMINATION OF BASE MODIFICATIONS OF NUCLEIC ACIDS
First Named Inventor/Applicant Name:	Yuk-Ming Dennis Lo
Customer Number:	20350
Filer:	David Dah-Wei Hsu/Heather Orosco
Filer Authorized By:	David Dah-Wei Hsu
Attorney Docket Number:	080015-028420US-1249377
Receipt Date:	19-JAN-2022
Filing Date:	19-JUL-2021
Time Stamp:	17:44:25
Application Type:	Utility under 35 USC 111(a)

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		1249377_Response_to_Restriction_Requirement.pdf	115934 0966cd880063dcc861794e37af2389d741600c7d	yes	14

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Response to Election / Restriction Filed	1	1
	Claims	2	12
	Applicant Arguments/Remarks Made in an Amendment	13	14
Warnings:			
Information:			
	Total Files Size (in bytes):	115934	

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.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. 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.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 17/379,544	Filing Date 07/19/2021	☐ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $100 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $480 =	
☐ 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).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	01/19/2022	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 30	Minus	** 30	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 4	Minus	*** 4	= 0		x $480 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	SLIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/NICHELE D PETERSON/
*** 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 number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	17/379,544
Filing Date	July 19, 2021
First Named Inventor	Yuk-Ming Dennis Lo
Art Unit	1642
Examiner Name	Laura B. Goddard
Attorney Docket Number	080015-1249377

U.S. PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code [1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1	8467976		Jun 18, 2013	Lo et al.	
	2	9116118		Aug 25, 2015	Turner et al.	
	3	9175338		Nov 3, 2015	Flusberg et al.	
	4	9175341		Nov 3, 2015	Flusberg et al.	
	5	10438691		Oct 8, 2019	Kim et al.	
	6	10465232		Nov 5, 2019	Wu et al.	
	7	10468121		Nov 5, 2019	Kermani et al.	
	8	10590484		Mar 17, 2020	Korlach et al.	
	9	11091794		Aug 17, 2021	Lo et al.	

U.S. PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code [1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	10	20110183320		Jul 28, 2011	Flusberg et al.	
	11	20130230909		Sep 5, 2013	Pan et al.	
	12	20140004511		Jan 2, 2014	Korlach et al.	
	13	20160017419		Jan 21, 2016	Chiu et al.	
	14	20170037464		Feb 9, 2017	Turner et al.	
	15	20170159132		Jun 8, 2017	Okino et al.	
	16	20170233802		Aug 17, 2017	Flusberg et al.	
	17	20180201993		Jul 19, 2018	Turner et al.	
	18	20190024162		Jan 24, 2019	Korlach et al.	

FOREIGN PATENT DOCUMENTS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T[5]
	19	2009013492	WO		Jan 29, 2009	The Chinese University of Hong Kong		
	20	2010027484	WO		Mar 11, 2010	Pacific Biosciences of California Inc.		
	21	2010068289	WO		Jun 17, 2010	Pacific Biosciences of California Inc.		
	22	2013170429	WO		Nov 21, 2013	BGI Genomics Co. Ltd.	English Abstract Attached	
	23	2013188846	WO		Dec 19, 2013	Harry Stylli		
	24	104053789	CN		Sep 17, 2014	BGI Genomics Co. Ltd.	English Abstract Attached	
	25	2014153757	WO		Oct 2, 2014	BGI Shenzhen	English Abstract Attached	
	26	104053789	CN		Feb 10, 2016	BGI Genomics Co. Ltd.	English Abstract Attached	
	27	2012380221	AU		Sep 29, 2016	BGI Genomics Co. Ltd.		
	28	2017012544	WO		Jan 26, 2017	The Chinese University of Hong Kong		
	29	101896167	KR		Sep 7, 2018	The Chinese University of Hong Kong	English Abstract of corresponding document: WO2009013492 (A1) Attached	

NON-PATENT LITERATURE DOCUMENTS

Examiner 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.	T[5]
	30	BEAULAURIER et al., Deciphering Bacterial Epigenomes Using Modern Sequencing Technologies, Nature Reviews Genetics, Volume 20 Issue 3, March 2019, pp. 157-172.	
	31	BLOW et al., The Epigenomic Landscape of Prokaryotes, PLOS Genetics, Volume 12, No. 2, February 12, 2016, pp. 1-28.	
	32	CLARK et al., Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing, Nucleic Acids Research, Volume 40, No. 4, Available online at: https://doi.org/10.1093/nar/gkr1146, 2012, 12 pages.	
	33	CLARK et al., Enhanced 5-Methylcytosine Detection in Single-Molecule, Real-Time Sequencing via Tet1 Oxidation, BMC Biology, Volume 11, No. 4, 2013, 10 pages.	

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

	34	EID et al., Real-Time DNA Sequencing from Single Polymerase Molecules, Science, Volume 323, No. 5910, January 2, 2009, pp. 133-138.	
	35	FENG et al., Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic, PLOS Computational Biology, Volume 9, No. 3, Available online at: https://doi.org/10.1371/journal.pcbi.1002935, March 2013, 10 pages.	
	36	FENG et al., qDNAmod: A Statistical Model-based Tool to Reveal Intercellular Heterogeneity of DNA Modification from SMRT Sequencing Data, Nucleic Acids Research, Volume 42, No. 22, November 17, 2014, pp. 13488-13499.	
	37	FLUSBERG et al., Direct Detection of DNA Methylation During Single-Molecule, Real-Time Sequencing, Nature Methods, Volume 7, No. 6, June 2010, pp. 461-465.	
	38	HE et al., TETI Promotes Leukemic Growth of AMLI-ETO+ AML Via 5-Hydroxymethylcytosine Marks and its Oncogenic Role and High Expression can be Antagonized Using Olaparib, an Inhibitor of its Binding Partner PARPI, Experimental Hematology, Volume 53, 2017, pp. S92-S93.	
	39	KILE et al., Correlation of Global and Gene-Specific DNA Methylation in Maternal-Infant Pairs, PLoS One, Volume 5, No. 10, October 29, 2010, 9 pages.	
	40	LIU et al., Bisulfite-free Direct Detection of 5-Methylcytosine and 5- Hydroxymethylcytosine at Base Resolution, Nature Biotechnology Letters, Volume 37, April 2019, pp. 424-429.	
	41	LIU et al., Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data, Nature Communications, Volume 10, No. 2449, June 4, 2019, 11 pages.	
	42	NI et al., Deepsignal: Detecting DNA Methylation State From Nanopore Sequencing Reads Using Deep-Learning, Bioinformatics, Volume 35, No. 22, 2019, pp. 4586-4595.	
	43	SCHADT et al., Modeling Kinetic Rate Variation in Third Generation DNA Sequencing Data to Detect Putative Modifications to DNA Bases, Genome Research, Volume 23, No. 1, January 2013, pp. 129-141.	
	44	SUGAI et al., Analysis of the DNA Methylation Level of Cancer-related Genes in Colorectal Cancer and the Surrounding Normal Mucosa, Clinical Epigenetics, Volume 9, No. 55, May 18, 2017, pp. 1-10.	
	45	SUZUKI et al., Agin: Measuring the Landscape of CpG Methylation of Individual Repetitive Elements, Bioinformatics, Volume 32, No.19, June 17, 2016, pp. 2911-2919.	
	46	SUZUKI, Observing Comprehensive DNA Methylomes via Single Molecule Real-Time Sequencing: Application to Diploid and Centromeric Methylation, Dissertation submitted to The Graduate School of Frontier Sciences, The University of Tokyo, December 13, 2017, 118 pages.	

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

EXAMINER SIGNATURE			
Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document.
[4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
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 contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	17/379,544
	Filing Date	July 19, 2021
	First Named Inventor	Yuk-Ming Dennis Lo
	Art Unit	1642
	Examiner Name	Laura B. Goddard
	Attorney Docket Number	080015-1249377

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e) (2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David D. Hsu/	Date (YYYY-MM-DD)	2021-12-22
Name/Print	David D. Hsu	Registration Number	71,659

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 January 2009 (29.01.2009)

PCT

(10) International Publication Number
WO 2009/013492 A1

(51) International Patent Classification:
G06F 19/00 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/GB2008/002524

(22) International Filing Date: 23 July 2008 (23.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/951,438 23 July 2007 (23.07.2007) US

(71) Applicant *(for all designated States except US)*: **THE CHINESE UNIVERSITY OF HONG KONG** [CN/CN]; Technology Licensing Office, Room 226, Pi Ch'iu Building, Shatin, New Territories (HK).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LO, Yuk-Ming Dennis** [GB/CN]; 4th Floor, 7 King Tak Street, Homantin, Kowloon (HK). **CHIU, Rossa Wai Kwun** [AU/CN]; Flat 1A, Block 1, Constellation Cove, 1 Hung Lam Drive, Tai Po, New Territories (HK). **CHAN, Kwan Chee** [CN/CN]; Flat A, 13/F, Block 34, Broadway Street, Mei Foo Sun Chuen, Kowloon (HK). **ZEE, Benny Chung Ying** [CA/CN]; Flat 18E, Tower 2, La Costa, Ma on Shan, New Territories (HK). **CHONG, Ka Chun** [CN/CN]; Flat 06, 29/F, Shin King House, Fu Shin Estate, Tai Po, New Territories (HK).

(74) Agents: **BUFTON, Karen** et al.; Mewburn Ellis LLP, 33 Gutter Lane, London EC2V 8AS (GB).

(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, BR, BW, BY, BZ, CA, CH, 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, PG, PH, PL, PT,

[Continued on next page]

(54) Title: DETERMINING A NUCLEIC ACID SEQUENCE IMBALANCE

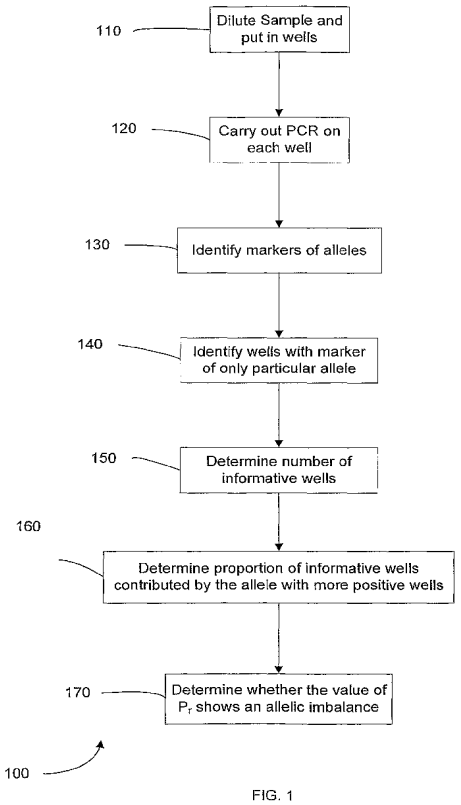


FIG. 1

(57) Abstract: Methods, systems, and apparatus are provided for determining whether a nucleic acid sequence imbalance exists within a biological sample. One or more cutoff values for determining an imbalance of, for example, the ratio of the two sequences (or sets of sequences) are chosen. The cutoff value may be determined based at least in part on the percentage of fetal DNA in a sample, such as maternal plasma, containing a background of maternal nucleic acid sequences. The cutoff value may also be determined based on an average concentration of a sequence per reaction. In one aspect, the cutoff value is determined from a proportion of informative wells that are estimated to contain a particular nucleic acid sequence, where the proportion is determined based on the above-mentioned percentage and/or average concentration. The cutoff value may be determined using many different types of methods, such as sequential probability ratio testing (SPRT).

WO 2009/013492 A1

RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, 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:
— *with international search report*

DETERMINING A NUCLEIC ACID SEQUENCE IMBALANCE

CLAIM OF PRIORITY

[0001] The present application claims priority from and is a non-provisional application of U.S. Provisional Application No. 60/951438, entitled "DETERMINING A NUCLEIC ACID SEQUENCE IMBALANCE" filed July 23, 2007 (Attorney Docket No. 016285-005200US), the entire contents of which are herein incorporated by reference for all purposes.

CROSS-REFERENCES TO RELATED APPLICATIONS

[0002] The present application is also related to concurrently filed non-provisional application entitled "DIAGNOSING FETAL CHROMOSOMAL ANEUPLOIDY USING GENOMIC SEQUENCING," (Attorney Docket No. 016285-005220US), the entire contents of which are herein incorporated by reference for all purposes.

FIELD OF THE INVENTION

[0003] This invention generally relates to the diagnostic testing of genotypes and diseases by determining an imbalance between two different nucleic acid sequences, and more particularly to the identification of Down syndrome, other chromosomal aneuploidies, mutations and genotypes in a fetus via testing a sample of maternal blood. The invention also relates to the detection of cancer, the monitoring of transplantation, and the monitoring of infectious diseases.

BACKGROUND

[0004] Genetic diseases, cancers, and other conditions often result from or produce an imbalance in two corresponding chromosomes or alleles or other nucleic acid sequences. That is an amount of one sequence relative to another sequence is larger or smaller than normal. Usually, the normal ratio is an even 50/50 ratio. Down Syndrome (trisomy 21) is such a disease having an imbalance of an extra chromosome 21.

[0005] Conventional prenatal diagnostic methods of trisomy 21 involve the sampling of fetal materials by invasive procedures such as amniocentesis or chorionic villus sampling, which pose a finite risk of fetal loss. Non-invasive procedures, such as screening by ultrasonography and biochemical markers, have been used to risk-stratify pregnant women prior to definitive invasive diagnostic procedures. However, these screening methods typically measure epiphenomena that are associated with trisomy 21 instead of the core

chromosomal abnormality, and thus have suboptimal diagnostic accuracy and other disadvantages, such as being highly influenced by gestational age.

[0006] The discovery of circulating cell-free fetal DNA in maternal plasma in 1997 offered new possibilities for noninvasive prenatal diagnosis (Lo, YMD and Chiu, RWK 2007 *Nat Rev Genet* 8, 71-77). While this method has been readily applied to the prenatal diagnosis of sex-linked (Costa, JM et al. 2002 *N Engl J Med* 346, 1502) and certain single gene disorders (Lo, YMD et al. 1998 *N Engl J Med* 339, 1734-1738), its application to the prenatal detection of fetal chromosomal aneuploidies has represented a considerable challenge (Lo, YMD and Chiu, RWK 2007, supra). First, fetal nucleic acids co-exist in maternal plasma with a high background of nucleic acids of maternal origin that can often interfere with the analysis (Lo, YMD et al. 1998 *Am J Hum Genet* 62, 768-775). Second, fetal nucleic acids circulate in maternal plasma predominantly in a cell-free form, making it difficult to derive dosage information of genes or chromosomes within the fetal genome.

[0007] Significant developments overcoming these challenges have recently been made (Benachi, A & Costa, JM 2007 *Lancet* 369, 440-442). One approach detects fetal-specific nucleic acids in the maternal plasma, thus overcoming the problem of maternal background interference (Lo, YMD and Chiu, RWK 2007, supra). Dosage of chromosome 21 was inferred from the ratios of polymorphic alleles in the placenta-derived DNA/RNA molecules. However, this method is less accurate when samples contain lower amount of the targeted gene and can only be applied to fetuses who are heterozygous for the targeted polymorphisms, which is only a subset of the population if one polymorphism is used.

[0008] Dhallan et al (Dhallan, R, *et al.* 2007, supra Dhallan, R, *et al.* 2007 *Lancet* 369, 474-481) described an alternative strategy of enriching the proportion of circulating fetal DNA by adding formaldehyde to maternal plasma. The proportion of chromosome 21 sequences contributed by the fetus in maternal plasma was determined by assessing the ratio of paternally-inherited fetal-specific alleles to non-fetal-specific alleles for single nucleotide polymorphisms (SNPs) on chromosome 21. SNP ratios were similarly computed for a reference chromosome. An imbalance of fetal chromosome 21 was then inferred by detecting a statistically significant difference between the SNP ratios for chromosome 21 and those of the reference chromosome, where significant is defined using a fixed p-value of ≤ 0.05. To ensure high population coverage, more than 500 SNPs were targeted per chromosome. However, there have been controversies regarding the effectiveness of formaldehyde to

enrich to a high proportion (Chung, GTY, et al. 2005 *Clin Chem* 51, 655-658), and thus the reproducibility of the method needs to be further evaluated. Also, as each fetus and mother would be informative for a different number of SNPs for each chromosome, the power of the statistical test for SNP ratio comparison would be variable from case to case (Lo, YMD & Chiu, RWK. 2007 *Lancet* 369, 1997). Furthermore, since these approaches depend on the detection of genetic polymorphisms, they are limited to fetuses heterozygous for these polymorphisms.

[0009] Using polymerase chain reaction (PCR) and DNA quantification of a chromosome 21 locus and a reference locus in amniocyte cultures obtained from trisomy 21 and euploid fetuses, Zimmermann et al (2002 *Clin Chem* 48, 362-363) were able to distinguish the two groups of fetuses based on the 1.5-fold increase in chromosome 21 DNA sequences in the former. Since a 2-fold difference in DNA template concentration constitutes a difference of only one threshold cycle (Ct), the discrimination of a 1.5-fold difference has been the limit of conventional real-time PCR. To achieve finer degrees of quantitative discrimination, alternative strategies are needed. Accordingly, some embodiments of the present invention use digital PCR (Vogelstein, B et al. 1999 *Proc Natl Acad Sci U S A* 96, 9236-9241) for this purpose.

[0010] Digital PCR has been developed for the detection of allelic ratio skewing in nucleic acid samples (Chang, HW *et al.* 2002 *J Natl Cancer Inst* 94, 1697-1703). Clinically, it has been shown to be useful for the detection of loss of heterozygosity (LOH) in tumor DNA samples (Zhou, W. *et al.* 2002 *Lancet* 359, 219-225). For the analysis of digital PCR results, sequential probability ratio testing (SPRT) has been adopted by previous studies to classify the experimental results as being suggestive of the presence of LOH in a sample or not (El Karoui at al. 2006 *Stat Med* 25, 3124-3133). In methods used in the previous studies, the cutoff value to determine LOH used a fixed reference ratio of the two alleles in the DNA of 2/3. As the amount, proportion, and concentration of fetal nucleic acids in maternal plasma are variable, these methods are not suitable for detecting trisomy 21 using fetal nucleic acids in a background of maternal nucleic acids in maternal plasma.

[0011] It is desirable to have a noninvasive test for fetal trisomy 21 (and other imbalances) detection based on circulating fetal nucleic acid analysis, especially one that is independent of the use of genetic polymorphisms and/or of fetal-specific markers. It is also desirable to have accurate determination of cutoff values and counting of sequences, which can reduce the

number of wells of data and/or the amount of maternal plasma nucleic acid molecules necessary for accuracy, thus providing increased efficiency and cost-effectiveness. It is also desirable that noninvasive tests have high sensitivity and specificity to minimize false diagnoses.

[0012] Another application for fetal DNA detection in maternal plasma is for the prenatal diagnosis of single gene disorders such as beta-thalassemia. However, as fetal DNA only constitutes a minor fraction of DNA in maternal plasma, it is thought that this approach can only detect a mutation that a fetus has inherited from its father, but which is absent from the mother. Examples of this include the 4 bp deletion in codon 41/42 of the beta-globin gene causing beta-thalassemia (Chiu RWK et al. 2002 *Lancet*, 360, 998-1000) and the Q890X mutation of the cystic fibrosis transmembrance conductance regulator gene causing cystic fibrosis (Gonzalez-Gonzalez et al 2002 *Prenat Diagn,* 22, 946-8). However, as both beta-thalassemia and cystic fibrosis are autosomal recessive conditions, in which the fetus would need to inherit a mutation from each parent before the disease would manifest itself, the detection of merely the paternally-inherited mutation would only increase the risk of having the fetus having the disease from 25% to 50%. Diagnostically this is not ideal. Thus, the main diagnostic application of the existing approach would be for the scenario when no paternally-inherited fetal mutation can be detected in maternal plasma, when the fetus can then be excluded from having the homozygous disease state. However, diagnostically, this approach has the disadvantage that the conclusion is made based on the negative detection of the paternal mutation. Thus, an approach which would allow the complete fetal genotype (be it homozygous normal, homozygous mutant, or heterozygous) to be determined from maternal plasma, without the above limitation, would be very desirable.

BRIEF SUMMARY

[0013] Embodiments of this invention provides methods, systems, and apparatus for determining whether a nucleic acid sequence imbalance (e.g., allelic imbalance, mutational imbalance, or chromosome imbalance) exists within a biological sample. One or more cutoff values for determining an imbalance of, for example, a ratio of amounts of the two sequences (or sets of sequences) are chosen.

[0014] In one embodiment, the cutoff value is determined based at least in part on the percentage of fetal (clinically relevant nucleic acid) sequences in a biological sample, such as maternal plasma or serum or urine, which contains a background of maternal nucleic acid

sequences. In another embodiment, the cutoff value is determined based on an average concentration of a sequence in a plurality of reactions. In one aspect, the cutoff value is determined from a proportion of informative wells that are estimated to contain a particular nucleic acid sequence, where the proportion is determined based on the above-mentioned percentage and/or average concentration.

[0015] The cutoff value may be determined using many different types of methods, such as SPRT, false discovery, confidence interval, receiver operating characteristic (ROC). This strategy further minimized the amount of testing required before confident classification could be made. This is of particular relevance to plasma nucleic acid analysis where the template amount is often limiting.

[0016] According to one exemplary embodiment, a method is provided for determining whether a nucleic acid sequence imbalance exists within a biological sample, the method comprising: receiving data from a plurality of reactions, wherein the data includes: (1) a first set of quantitative data indicating a first amount of a clinically relevant nucleic acid sequence; and (2) a second set of quantitative data indicating a second amount of a background nucleic acid sequence different from the clinically relevant nucleic acid sequence; determining a parameter from the two data sets; deriving a first cutoff value from an average concentration of a reference nucleic acid sequence in each of the plurality of reactions, wherein the reference nucleic acid sequence is either the clinically relevant nucleic acid sequence or the background nucleic acid sequence; comparing the parameter to the first cutoff value; and based on the comparison, determining a classification of whether a nucleic acid sequence imbalance exists.

[0017] According to another exemplary embodiment, a method is provided for determining whether a nucleic acid sequence imbalance exists within a biological sample, the method comprising: receiving data from a plurality of reactions, wherein the data includes: (1) a first set of quantitative data indicating a first amount of a clinically relevant nucleic acid sequence; and (2) a second set of quantitative data indicating a second amount of a background nucleic acid sequence different from the clinically relevant nucleic acid sequence, wherein the clinically relevant nucleic acid sequence and the background nucleic acid sequence come from a first type of cells and from one or more second types of cells; determining a parameter from the two data sets; deriving a first cutoff value from a first percentage resulting from a measurement of an amount of a nucleic acid sequence from the first type of cells in the

biological sample; comparing the parameter to the cutoff value; and based on the comparison, determining a classification of whether a nucleic acid sequence imbalance exists.

[0018] Other embodiments of the invention are directed to systems and computer readable media associated with methods described herein.

[0019] A better understanding of the nature and advantages of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] FIG. 1 is a flowchart illustrating a digital PCR experiment.

[0021] FIG. 2A illustrates a digital RNA-SNP and RCD method according to an embodiment of the present invention.

[0022] FIG. 2B shows a table of examples of frequently detectable chromosomal aberrations in cancers.

[0023] FIG. 3 illustrates a graph having SPRT curves used to determine Down syndrome according to an embodiment of the present invention.

[0024] FIG. 4 shows a method of determining a disease state using a percentage of fetal cells according to an embodiment of the present invention.

[0025] FIG. 5 shows a method of determining a disease state using an average concentration according to an embodiment of the present invention.

[0026] FIG. 6 shows a table that tabulates the expected digital RNA-SNP allelic ratio and P_r of trisomy 21 samples for a range of template concentrations expressed as the average reference template concentration per well (m_r) according to an embodiment of the present invention.

[0027] FIG. 7 shows a table that tabulates the expected P_r for the fractional fetal DNA concentrations of 10%, 25%, 50% and 100% in trisomy 21 samples at a range of template concentrations expressed as the average reference template concentration per well (m_r) according to an embodiment of the present invention.

[0028] FIG. 8 shows a plot illustrating the degree of differences in the SPRT curves for m_r values of 0.1, 0.5 and 1.0 for digital RNA-SNP analysis according to an embodiment of the present invention.

[0029] FIG. 9A shows a table of a comparison of the effectiveness of the new and old SPRT algorithms for classifying euploid and trisomy 21 cases in 96-well digital RNA-SNP analyses according to an embodiment of the present invention.

[0030] FIG. 9B shows a table of a comparison of the effectiveness of the new and old SPRT algorithms for classifying euploid and trisomy 21 cases in 384-well digital RNA-SNP analyses according to an embodiment of the present invention.

[0031] FIG. 10 is a table showing the percentages of fetuses correctly and incorrectly classified as euploid or aneuploid and those not classifiable for the given informative counts according to an embodiment of the present invention.

[0032] FIG. 11 is a table 1100 showing computer simulations for digital RCD analysis for a pure (100%) fetal DNA sample according to an embodiment of the present invention.

[0033] FIG. 12 is a table 1200 showing results of computer simulation of accuracies of digital RCD analysis at m_r=0.5 for the classification of samples from euploid or trisomy 21 fetuses with different fractional concentrations of fetal DNA according to an embodiment of the present invention.

[0034] FIG. 13A shows a table 1300 of digital RNA-SNP analysis in placental tissues of euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

[0035] FIG. 13B shows a table 1350 of digital RNA-SNP analysis of maternal plasma from euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

[0036] FIG. 14A-14C show plots illustrating a cutoff curve resulting from an RCD analysis according to an embodiment of the present invention.

[0037] FIG. 15A shows a table of digital RNA-SNP analysis in placental tissues of euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

[0038] FIG. 15B shows a table of digital RNA-SNP data of the 12 reaction panels from one maternal plasma sample according to an embodiment of the present invention.

[0039] FIG. 15C shows a table of digital RNA-SNP analysis of maternal plasma from euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

[0040] FIG. 16A shows a table for a digital RNA-SNP analysis of euploid and trisomy 18 placentas according to an embodiment of the present invention.

[0041] FIG. 16B shows an SPRT interpretation of digital RNA-SNP data for euploid and trisomy 18 placentas according to an embodiment of the present invention.

[0042] FIG. 17 shows a table of a digital RCD analysis of 50% placental/maternal blood cell DNA mixtures of euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

[0043] FIG. 18 shows a SPRT curve illustrating the decision boundaries for correct classification according to an embodiment of the present invention.

[0044] FIG. 19 shows a table of digital RCD analysis of amniotic fluid samples from euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

[0045] FIG. 20 shows a table of digital RCD analysis of placental DNA samples from euploid and trisomy 18 pregnancies (E= euploid; T18= trisomy 18) according to an embodiment of the present invention.

[0046] FIG. 21 shows a table of a multiplex digital RCD analysis of 50% placental/maternal blood cell DNA mixtures of euploid and trisomy 21 pregnancies (E= euploid; T21= trisomy 21; U= unclassified) according to an embodiment of the present invention.

[0047] FIGS. 22A and 22B show a table of a multiplex digital RCD analysis of 50% euploid or trisomy 21 placental genomic DNA/50% maternal buffy coat DNA mix according to an embodiment of the present invention. Unclass denotes unclassifiable and T21 denotes trisomy 21.

[0048] FIG. 23 shows a scenario where both the male and female partners carry the same mutation.

[0049] FIG. 24A shows a table of a digital RMD analysis of female/male and male/male DNA mixtures according to an embodiment of the present invention.

[0050] FIG. 24B shows a table of a digital RMD analysis of mixtures with 25% female and 75% male DNA according to an embodiment of the present invention.

[0051] FIG. 25 shows a table of a digital RMD analysis of 15%-50% DNA mixtures mimicking maternal plasma samples for HbE mutation according to an embodiment of the present invention.

[0052] FIG. 26A shows a table of a digital RMD analysis of 5%-50% DNA mixtures mimicking maternal plasma samples for CD41/42 mutation according to an embodiment of the present invention.

[0053] FIG. 26B shows a table of a digital RMD analysis of 20% DNA mixtures mimicking maternal plasma samples for CD41/42 mutation according to an embodiment of the present invention.

[0054] FIG. 27 shows a block diagram of an exemplary computer apparatus usable with system and methods according to embodiments of the present invention.

DEFINITIONS

[0055] The term "*biological sample*" as used herein refers to any sample that is taken from a subject (*e.g.*, a human, such as a pregnant woman) and contains one or more nucleic acid molecule(s)of interest.

[0056] The term "*nucleic acid*" or "*polynucleotide*" refers to a deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) and a polymer thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions), alleles, orthologs, SNPs, and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* **19**:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* **260**:2605-2608 (1985); and Rossolini *et al.*, *Mol. Cell. Probes* **8**:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, small noncoding RNA, micro RNA (miRNA), Piwi-interacting RNA, and short hairpin RNA (shRNA) encoded by a gene or locus.

[0057] The term "*gene*" means the segment of DNA involved in producing a polypeptide chain. It may include regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[0058] The term "*reaction*" as used herein refers to any process involving a chemical, enzymatic, or physical action that is indicative of the presence or absence of a particular polynucleotide sequence of interest. An example of a "*reaction*" is an amplification reaction such as a polymerase chain reaction (PCR). Another example of a "*reaction*" is a sequencing reaction, either by synthesis or by ligation. An "*informative reaction*" is one that indicates the presence of one or more particular polynucleotide sequence of interest, and in one case where only one sequence of interest is present. The term "*well*" as used herein refers to a reaction at a predetermined location within a confined structure, e.g., a well-shaped vial, cell, or chamber in a PCR array.

[0059] The term "*clinically relevant nucleic acid sequence*" as used herein can refer to a polynucleotide sequence corresponding to a segment of a larger genomic sequence whose potential imbalance is being tested or to the larger genomic sequence itself. One example is the sequence of chromosome 21. Other examples include chromosome 18, 13, X and Y. Yet other examples include mutated genetic sequences or genetic polymorphisms or copy number variations that a fetus may inherit from one or both of its parents. Yet other examples include sequences which are mutated, deleted, or amplified in a malignant tumor, e.g. sequences in which loss of heterozygosity or gene duplication occur. In some embodiments, multiple clinically relevant nucleic acid sequences, or equivalently multiple makers of the clinically relevant nucleic acid sequence, can be used to provide data for detecting the imbalance. For instance, data from five non-consecutive sequences on chromosome 21 can be used in an additive fashion for the determination of a possible chromosomal 21 imbalance, effectively reducing the need of sample volume to 1/5.

[0060] The term "*background nucleic acid sequence*" as used herein refers to a nucleic acid sequence whose normal ratio to the clinically relevant nucleic acid sequence is known, for instance a 1-to-1 ratio. As one example, the background nucleic acid sequence and the clinically relevant nucleic acid sequence are two alleles from the same chromosome that are distinct due to heterozygosity. In another example, the background nucleic acid sequence is one allele that is heterozygous to another allele that is the clinically relevant nucleic acid sequence. Moreover, some of each of the background nucleic acid sequence and the clinically relevant nucleic acid sequence may come from different individuals.

[0061] The term "*reference nucleic acid sequence*" as used herein refers to a nucleic acid sequence whose average concentration per reaction is known or equivalently has been measured.

[0062] The term "*overrepresented nucleic acid sequence*" as used herein refers to the nucleic acid sequence among two sequences of interest (e.g., a clinically relevant sequence and a background sequence) that is in more abundance than the other sequence in a biological sample.

[0063] The term "*based on*" as used herein means "*based at least in part on*" and refers to one value (or result) being used in the determination of another value, such as occurs in the relationship of an input of a method and the output of that method. The term "*derive*" as used herein also refers to the relationship of an input of a method and the output of that method, such as occurs when the derivation is the calculation of a formula.

[0064] The term "*quantitative data*" as used herein means data that are obtained from one or more reactions and that provide one or more numerical values. For example, the number of wells that show a fluorescent marker for a particular sequence would be quantitative data.

[0065] The term "*parameter*" as used herein means a numerical value that characterizes a quantitative data set and/or a numerical relationship between quantitative data sets. For example, a ratio (or function of a ratio) between a first amount of a first nucleic acid sequence and a second amount of a second nucleic acid sequence is a parameter.

[0066] The term "*cutoff value*" as used herein means a numerical value whose value is used to arbitrate between two or more states (e.g. diseased and non-diseased) of classification for a biological sample. For example, if a parameter is greater than the cutoff value, a first classification of the quantitative data is made (e.g. diseased state); or if the parameter is less than the cutoff value, a different classification of the quantitative data is made (e.g. non-diseased state).

[0067] The term "*imbalance*" as used herein means any significant deviation as defined by at least one cutoff value in a quantity of the clinically relevant nucleic acid sequence from a reference quantity. For example, the reference quantity could be a ratio of 3/5, and thus an imbalance would occur if the measured ratio is 1:1.

DETAILED DESCRIPTION OF THE INVENTION

[0068] This invention provides methods, systems, and apparatus for determining whether an increase or decrease compared to a reference (e.g. non-diseased) quantity of a clinically-relevant nucleic acid sequence in relation to other non-clinically-relevant sequences (e.g., a chromosomal or allelic imbalance) exists within a biological sample. One or more cutoff values are chosen for determining whether a change compared to the reference quantity exists (i.e. an imbalance), for example, with regards to the ratio of amounts of two sequences (or sets of sequences). The change detected in the reference quantity may be any deviation (upwards or downwards) in the relation of the clinically-relevant nucleic acid sequence to the other non-clinically-relevant sequences. Thus, the reference state may be any ratio or other quantity (e.g. other than a 1-1 correspondence), and a measured state signifying a change may be any ratio or other quantity that differs from the reference quantity as determined by the one or more cutoff values.

[0069] The clinically relevant nucleic acid sequence and the background nucleic acid sequence may come from a first type of cells and from one or more second types of cells. For example, fetal nucleic acid sequences originating from fetal/placental cells are present in a biological sample, such as maternal plasma, which contains a background of maternal nucleic acid sequences originating from maternal cells. Thus, in one embodiment, the cutoff value is determined based at least in part on a percentage of the first type of cells in a biological sample. Note the percentage of fetal sequences in a sample may be determined by any fetal-derived loci and not limited to measuring the clinically-relevant nucleic acid sequences. In another embodiment, the cutoff value is determined at least in part on the percentage of tumor sequences in a biological sample, such as plasma, serum, saliva or urine, which contains a background of nucleic acid sequences derived from the non-malignant cells within the body.

[0070] In yet another embodiment, the cutoff value is determined based on an average concentration of a sequence in a plurality of reactions. In one aspect, the cutoff value is determined from a proportion of informative wells that are estimated to contain a particular nucleic acid sequence, where the proportion is determined based on the above-mentioned percentage and/or average concentration. The cutoff value may be determined using many different types of methods, such as SPRT, false discovery, confidence interval, receiver operating characteristic (ROC). This strategy further minimizes the amount of testing required before confident classification can be made. This is of particular relevance to

plasma nucleic acid analysis where the template amount is often limiting. Although presented with respect to digital PCR, other methods may be used.

[0071] Digital PCR involves multiple PCR analyses on extremely dilute nucleic acids such that most positive amplifications reflect the signal from a single template molecule. Digital PCR thereby permits the counting of individual template molecules. The proportion of positive amplifications among the total number of PCRs analyzed allows an estimation of the template concentration in the original or non-diluted sample. This technique has been proposed to allow the detection of a variety of genetic phenomena (Vogelstein, B et al. 1999, supra) and has previously been used for the detection of loss of heterozygosity in tumor samples (Zhou, W. *et al.* 2002, supra) and in the plasma of cancer patients (Chang, HW *et al.* 2002, supra). Since template molecule quantification by digital PCR does not rely on dose-response relationships between reporter dyes and nucleic acid concentrations, its analytical precision should theoretically be superior to that of real-time PCR. Hence, digital PCR could potentially allow the discrimination of finer degrees of quantitative differences between target and reference loci.

[0072] To test this, we first assessed if digital PCR could determine the allelic ratio of *PLAC4* mRNA (Lo, YMD, *et al.* 2007 *Nat Med* 13, 218-223), a placental transcript from chromosome 21, in maternal plasma and thereby distinguish trisomy 21 and euploid fetuses. This approach is referred as the digital RNA-SNP method. We then evaluated whether the increased precision of digital PCR would allow the detection of fetal chromosomal aneuploidies without depending on genetic polymorphisms. We call this digital relative chromosome dosage (RCD) analysis. The former approach is polymorphism-dependent but requires less precision in quantitative discrimination while the latter approach is polymorphism-independent but requires a higher precision for quantitative discrimination.

I. DIGITAL RNA-SNP

A. Overview

[0073] Digital PCR is capable of detecting the presence of allelic ratio skewing of two alleles in a DNA sample. For example, it has been used to detect loss of heterozygosity (LOH) in a tumor DNA sample. Assuming that there are two alleles in the DNA sample, namely A and G, and the A allele would be lost in the cells with LOH. When LOH is present

in 50% of cells in the tumor sample, the allelic ratio of G:A in the DNA sample would be 2:1. However, if LOH is not present in the tumor sample, the allelic ratio of G:A would be 1:1.

[0074] FIG. 1 is a flowchart 100 illustrating a digital PCR experiment. In step 110, the DNA sample is diluted and then distributed to separate wells. Note that the inventors have determined that some plasma nucleic acid species are already quite diluted in the original sample. Accordingly, there is no need for dilution for some templates, if they are already present at the necessary concentrations. In the previous studies (e.g. Zhou et al 2002, supra), a DNA sample is diluted to an extent such that the average concentration of a specific "template DNA" is approximately 0.5 molecule of one of the two templates per well. Note that the term "template DNA" appears to refer to either the A or the G alleles, and that there is no rationale provided for this specific concentration.

[0075] In step 120, in each well, a PCR process is carried out to detect the A and/or the G allele simultaneously. In step 130, the markers in each well are identified (e.g. via fluorescence), e.g. A, G, A and G, or neither. In the absence of LOH, the abundance of the A and the G alleles in the DNA sample would be the same (one copy each per cell). Therefore, the probabilities of a well being positive for the A allele and for the G allele would be the same. This would be reflected by the similar numbers of wells being positive for the A or the G alleles. However, when LOH is present in 50% or greater of cells in a tumor sample, the allelic ratio of the G and the A alleles would be at least 2:1. Previous methods simply assumed that the sample was at least 50% cancerous. Thus, the probability of a well being positive for the G allele would be higher than that for the A allele. As a result, the number of wells being positive for the G allele would be higher than that for the A allele.

[0076] In step 140, to classify the digital PCR results, the number of wells being positive for each allele but not the other would be counted. In the above example, the number of wells being positive for the A allele but negative for the G allele, and the number of wells positive for the G allele but negative for the A allele are counted. In one embodiment, the allele showing less positive wells is regarded as the reference allele.

[0077] In step 150, the total number of informative wells is determined as the sum of the numbers of positive wells for either of the two alleles. In step 160, the proportion (P_r) of informative wells (an example of a parameter) contributed by the allele with more positive wells is calculated.

P_r = No. of wells only positive for the allele with more positive wells / Total no. of wells positive for only one allele (A or G).

Other embodiments could use all wells with one of the alleles divided by all wells with at least one allele.

[0078] In step 170, it is determined whether the value of P_r shows an allelic imbalance. As accuracy and efficiency are desired, this task is not straightforward. One method for determining an imbalance uses a Bayesian-type likelihood method, sequential probability ratio testing (SPRT). SPRT is a method which allows two probabilistic hypotheses to be compared as data accumulate. In other words, it is a statistical method to classify the results of digital PCR as being suggestive of the presence or absence of allelic skewing. It has the advantage of minimizing the number of wells to be analyzed to achieve a given statistical power and accuracy.

[0079] In an exemplary SPRT analysis, the experimental results would be tested against the null and alternative hypotheses. The alternative hypothesis is accepted when there is allelic ratio skewing in the sample. The null hypothesis is accepted when there is no allelic ratio skewing in the sample. The value P_r would be compared with two cutoff values to accept the null or alternative hypotheses. If neither hypothesis is accepted, the sample would be marked as unclassified which means that the observed digital PCR result is not sufficient to classify the sample with the desired statistical confidence.

[0080] The cutoff values for accepting the null or alternative hypotheses have typically been calculated based on a fixed value of P_r under the assumptions stated in the hypotheses. In the null hypothesis, the sample is assumed to exhibit no allelic ratio skewing. Therefore, the probabilities of each well being positive for the A and the G alleles would be the same and, hence, the expected value of P_r would be 1/2. In the alternative hypothesis, the expected value of P_r has been taken to be 2/3 or about halfway between 0.5 and 2/3, e.g. 0.585. Also, due to a limited number of experiments, one can choose an upper bound (.585+3/N) and a lower bound taken as (.585-3/N).

B. Detection of Down Syndrome

[0081] In one embodiment of the present invention, digital SNP is used to detect fetal Down syndrome from a pregnant woman's plasma. Using markers specific to fetal/placental cells, the ratio of alleles in chromosome 21 may be measured. For example, to determine if

an observed degree of overrepresentation of a *PLAC4* allele is statistically significant, SPRT is used.

[0082] According to one exemplary embodiment, digital RNA-SNP determines an imbalance in the ratio of polymorphic alleles of an A/G SNP, rs8130833, located on *PLAC4* mRNA which is transcribed from chromosome 21 and expressed by the placenta. For a heterozygous euploid fetus, the A and G alleles should be equally represented in the fetal genome (1:1 genomic ratio); while in trisomy 21, the trisomic chromosome 21 would be associated with an additional copy of one of the SNP alleles in the fetal genome giving a 2:1 ratio. The aim of the digital PCR analysis is to determine whether the amounts of the two *PLAC4* alleles in the analyzed sample are equal or otherwise. Thus, both the A and G *PLAC4* alleles are the target templates. A real-time PCR assay was designed to amplify *PLAC4* mRNA and the two SNP alleles were discriminated by TaqMan fluorescent probes. A schematic illustration of the analytical steps is shown in Fig. 2A.

[0083] FIG. 2A illustrates a digital RNA-SNP method 200 according to an embodiment of the present invention. In step 210, the sample is received. In step 220, the nucleic acid sequence, e.g. *PLAC4* mRNA, is quantified in extracted RNA samples. In one embodiment, this is done by real-time PCR for *PLAC4* mRNA. In one aspect, this step provides the operator with an idea about how much dilution is required before the target reaches the 'realm' of digital PCR analysis.

[0084] In step 230, the sample is diluted. In step 240, a concentration of the diluted sample is measured. The diluted sample concentration may be confirmed to be ~1 template/well (i.e. reference or non-reference sequence or either allele). Some embodiments use techniques described in section IV for this measurement. For example, we distributed the diluted sample to 96 wells for real-time PCR analysis to confirm that a usable dilution has been achieved. The dilution concentration may also be left as an unknown, thus removing this step as will be explained later.

[0085] In step 250, digital PCR is performed on each well of the array. For example, the same diluted sample was distributed to 384-wells for real-time PCR analysis. From the PCR results, an amount of markers for each nucleic acid sequence and the number of informative wells is identified. An informative well is defined as one that is only positive for the A or G allele but not both. In step 260, a calculation of an expected value of P_r is performed. These steps will be described in more detail later. The calculation includes determining a parameter

from values determined in step 250. For example, the actual average template concentration per well may be calculated.

[0086] In step 270, an SPRT or other likelihood-ratio test may be performed to determine whether or not an imbalance exists. For a euploid case, we expect an equal number of A-positive and G-positive wells. However, when template molecules from a trisomy 21 fetus are analyzed, the number of wells containing just one allele should be higher than that containing just the other allele. In short, allelic imbalance is expected for trisomy 21.

[0087] As mentioned above, SPRT is a Bayesian-type likelihood method which allows two probabilistic hypotheses to be compared as data accumulate. In digital PCR analysis for trisomy 21 detection, the alternative hypothesis is accepted when allelic imbalance exists (i.e. trisomy 21 detected); and the null hypothesis is accepted when there is no allelic imbalance (i.e. trisomy 21 not detected). The allele with the higher number of counts is referred as the potentially overrepresented allele and its proportion among all informative wells (P_r) would be calculated. SPRT is applied to determine if the P_r indicates sufficient degree of allelic imbalance that would be expected for a trisomy 21 sample.

[0088] Operationally, SPRT can be applied and interpreted through the use of graphs with a pair of SPRT curves that are constructed to define the probabilistic boundaries for accepting or rejecting either hypothesis. FIG. 3 illustrates a graph having SPRT curves for determining Down syndrome according to an embodiment of the present invention. The SPRT curves plot the required proportion of informative wells positive for the potentially overrepresented allele P_r, (y-axis) for a given total number of informative wells (x-axis) when confident classification could be made. As depicted in Fig. 3, the upper curve sets the probabilistic boundaries for accepting the alternative hypothesis; while the lower curve sets the probabilistic boundaries for accepting the null hypothesis.

[0089] The experimentally derived P_r value would be compared with the expected value of P_r to accept or reject either hypothesis. If the null hypothesis was accepted, the samples were classified as having been obtained from pregnant women with euploid fetuses. If the alternative hypothesis was accepted, the samples were classified as having been obtained from pregnant women with trisomy 21 fetuses. Alternatively, either hypothesis could not be accepted if the P_r for the given number of informative counts has not yet reached the required level of statistical confidence for disease classification. These cases were deemed unclassifiable until more data were available. If disease classification is not possible,

additional 384-well plates may be performed until the aggregated data become classifiable by SPRT.

[0090] SPRT thus offers the advantage that a smaller amount of testing is required for a given level of confidence than other statistical methods. In practical terms, SPRT allows the acceptance or rejection of either of the hypotheses as soon as the required amount of data has been accumulated and thus minimizes unnecessary additional analyses. This feature is of particular relevance to the analysis of plasma nucleic acids which are generally present at low concentrations where the number of available template molecules is limiting. In addition to a strict classification, the classification may also include a percent accuracy. For example, a classification resulting from a comparison with a cutoff value may provide that a sample shows a likelihood of a nucleic acid sequence imbalance with a certain percentage, or equivalently that determined imbalance is accurate to a certain percentage or other value.

[0091] A similar approach could be applied to determine the genotype of a fetus with regard to either a mutation or genetic polymorphism, using fetal nucleic acids in maternal plasma or serum. One should recall that a fetus would inherit half of its genome from its mother. As an illustration, consider a particular genetic locus with two alleles, A and B. If the mother is a heterozygote with a genotype of AB, the fetus could theoretically have a genotype of AA, BB or AB. If the fetus has a genotype of AB, i.e., the same as that of the mother, then there will only be nucleic acids of the AB genotype (from both the mother and fetus) in maternal plasma. Thus, nucleic acid or allelic balance is seen in maternal plasma. On the other hand, if the fetus has a genotype of AA or BB, then there would be allelic imbalance with an overrepresentation of the A or the B allele, respectively, in maternal plasma. This consideration is also applicable to disease-causing mutations (e.g. those causing cystic fibrosis or beta-thalassemia or spinal muscular atrophy), in which case A could be considered as the wildtype allele and B could be considered as the mutant allele.

II. DIGITAL RCD

[0092] A disadvantage of digital RNA-SNP is that it can only be applied to cases heterozygous for the analyzed SNP. One improvement is that it would be ideal if a noninvasive test for detecting fetal trisomy 21 or other fetal chromosomal aneuploidies (e.g. trisomy 18, 13, and the sex chromosome aneuploidies) based on circulating fetal nucleic acid analysis were independent of the use of genetic polymorphisms. Thus, in one embodiment, chromosome dosage is determined by digital PCR analysis of a non-polymorphic

chromosome 21 locus relative to one located on a reference chromosome, namely chromosome 1 in this study. A change of the ratio of chromosome 21 to chromosome 1 from 2:2 in the genome of a euploid fetus is differentiated from a trisomy 21 case. In digital PCR analysis for trisomy 21 detection, the two hypotheses to be compared would be the null hypothesis that there is no chromosomal imbalance (i.e. trisomy 21 not detected) and the alternative hypothesis that a chromosomal imbalance exists (i.e. trisomy 21 detected).

[0093] This approach can be generalized to the other chromosomes involved in other chromosomal aneuploidies, e.g. chromosome 18 in trisomy 18, chromosome 13 in trisomy 13, chromosome X in Turner syndrome. In addition, apart from chromosome 1, other chromosomes not involved in the aneuploidies concerned can also be used as a reference chromosome. A similar approach can also be applied to the detection of cancer, by analyzing the change of ratio of a chromosome commonly deleted, in part, in cancer, to a reference chromosome. Examples of the former include chromosome 5q in colorectal cancer, chromosome 3p in lung cancer and chromosome 9p in nasopharyngeal carcinoma. FIG. 2B lists some common cancer-related chromosomal aberrations which result in sequence imbalance.

[0094] FIG. 2A also illustrates a digital RCD method 205 according to an embodiment of the present invention. In one embodiment for steps 220-230, extracted DNA is quantified, for example, via Nanodrop techniques and diluted to a concentration of approximately one target template from either chromosomes 21 or the normalizing chromosome (such as chromosome 1) per well. In one embodiment in step 240, the confirmation may be performed by analyzing the diluted DNA sample by the assay using the chromosome 1 probe only in a 96-well format to confirm if ~37% level of the wells were negative before proceeding to digital RCD analysis using both TaqMan probes in 384-well plates. The significance of the 37% will be described later in Section IV.

[0095] The testing of step 240 and results of step 250 may be done with a real-time PCR assay designed to amplify a paralogous sequence (Deutsch, S. et al. 2004 *J Med Genet* 41, 908-915) present on both chromosomes which are distinguished by paralogous sequence variations that are discriminated by a pair of TaqMan probes. In this context, an informative well is defined as one that is positive for either the chromosome 21 or chromosome 1 locus but not both. For a euploid fetus, the number of informative wells positive for either locus should be approximately equal. For a trisomy 21 fetus, there should be an overrepresentation

of wells positive for chromosome 21 than chromosome 1. The exact proportion of the overrepresentation is described in the following sections.

III. INCORPORATING PERCENTAGE OF FETAL SEQUENCES

[0096] A disadvantage of embodiments of methods 200 and 205 described above is that fetal specific markers are required. Accordingly, in one embodiment of the present invention, non-fetal specific markers are used. In order to use such non-fetal specific markers, embodiments of the present invention measure the fractional concentration of fetal DNA in the maternal plasma (i.e. the biological sample). With such information, a more useful value of P_r may be calculated as follows.

[0097] Even with the small fractional percentage of fetal DNA in maternal plasma, a trisomy 21 fetus would contribute an additional dose of chromosome 21 sequences per genome-equivalent (GE) of fetal DNA released into maternal plasma. For example, a maternal plasma sample from a euploid pregnancy containing 50 GE/ml of total DNA with 5 GE/ml DNA contributed by the fetus (i.e. 10% fetal DNA fractional concentration) should contain a total of 100 copies (90 maternal copies + 10 fetal copies) of chromosome 21 sequences per milliliter of maternal plasma. For a trisomy 21 pregnancy, each fetal GE would contribute three copies of chromosome 21, resulting in a total of 105 copies/ml (90 maternal copies + 15 fetal copies) of chromosome 21 sequences in maternal plasma. At 10% fetal DNA concentration, the amount of chromosome 21 derived sequences in the maternal plasma of a trisomic pregnancy would therefore be 1.05 times that of a euploid case. Thus, if an analytical approach could be developed to determine this small degree of quantitative difference, a polymorphism-independent test for noninvasive prenatal diagnosis of fetal trisomy 21 would be achieved.

[0098] Accordingly, the degree of overrepresentation would be dependent on the fractional fetal DNA concentration in the analyzed DNA sample. For example, when placental DNA is analyzed, the theoretical RCD ratio in the fetal genome should be 3:2, i.e. 1.5-fold difference. However, as described above, the theoretical RCD ratio would decrease to 1.05 when a maternal plasma sample containing 10% fetal DNA is analyzed. The experimentally derived P_r is calculated by dividing the number of wells positive only for the chromosome 21 locus by the total number of informative wells. The experimentally derived P_r is subjected to the SPRT analysis with the calculated P_r and the theoretical RCD ratio.

[0099] FIG. 4 shows a method 400 of determining a disease state using a percentage of fetal nucleic acids according to an embodiment of the present invention. In step 410, the fractional percentage of fetal material is measured. In one embodiment, the fractional percentage is determined by measuring the amount of a fetal-specific marker (e.g. Y-chromosome, genetic polymorphism markers (e.g. SNPs), placental epigenetic signatures) in relation to a non-fetal-specific marker (i.e. gene sequence present in both mother and fetus). The actual measurement could be done by real-time PCR, digital PCR, sequencing reactions (including massively parallel genomic sequencing) or any other quantitative methods. In one aspect, it is preferable not to use the gene target that could potentially be in allelic imbalance for this measurement.

[0100] In step 420, digital PCR or other measurement method is performed, including diluting the sample, placing the diluted sample in the wells, and measuring the reactions in each well. In step 430, the PCR results are used to identify markers of different reference nucleic acid sequences (such as chromosomes or alleles). In step 440, the actual proportion (P_r) of the overrepresented sequence is calculated. In step 450, the cutoff value for determining a disease state is calculated using the percentage of fetal material in the sample. In step 460, from the actual P_r and the cutoff value, it is determined whether an imbalance exists.

[0101] In one embodiment, the fractional percentage of reference nucleic acid sequences is incorporated in a digital RNA-SNP method. Thus, when investigating a LOH due to cancer cells, this can be done with tumor samples with less than 50% cancer cells. It also may be used on samples with greater than 50% cancer cells to obtain a more accurate P_r and thus reduce the number of false positives, which would lead to incorrect diagnoses. In another embodiment, the fetal nucleic acid percentage is incorporated in a digital PCR method to determine if a fetus has inherited a parental gene mutation (e.g. that causing cystic fibrosis or beta-thalassemia or spinal muscular atrophy) or polymorphism from maternal plasma nucleic acid analysis.

IV. INCORPORATING AVERAGE CONCENTRATION PER WELL

[0102] Another disadvantage of previous methods (e.g. Zhou, W. *et al.* 2002, supra) is that the average concentration of templates per well (m) is required to be 1 per well. Given that it is difficult to obtain an exact concentration, this can lead to inaccuracies. Furthermore, even with an exact concentration of 1 template per well, previous methods have ignored the

statistical distribution of the templates in a well. In the previous methods, i.e. the old algorithm, the expected value of P_r for accepting the alternative hypothesis is assumed to be the allelic ratio and, thus, is independent of the average concentration of the template DNA per well.

[0103] However, due to a natural statistical variation of the templates in the diluted sample, there will not be exactly 1 template per well. Embodiments of the present invention measure the average concentration of at least one of the sequences, which is then used to calculate the cutoff value, i.e. the expected P_r. In one aspect, this calculation involves a statistical distribution to determine a probability of a well containing the different nucleic acid sequences, which is then used to determine the expected P_r.

[0104] In one embodiment, the average concentration is taken of one reference nucleic acid sequence, which in one instance is nucleic acid sequence with the lower concentration in the DNA sample. In the case of a sample without an imbalance, the concentrations of the two sequences in the sample would be the same and either one can be regarded as the reference allele. In the case of a sample with, for example, LOH, the allele which is deleted in the cancer cells would be regarded as the reference allele. The average concentration of the reference allele would be denoted as m_r. In another embodiment, the sequence with the higher concentration may be taken as the reference sequence.

A. Digital-SNP. Example using SPRT and digital PCR

[0105] FIG. 5 shows a method 500 of determining a disease state using an average template concentration according to an embodiment of the present invention. In step 510, an amount of the different sequences are measured. This may be done for example by counting the makers in a digital PCR experiment as explained above. However, it may be done by other methods that do not involve an amplification step or that do not use a fluorescent marker, but could use other properties, such as physical properties like mass, specific optical properties or base-pairing properties.

[0106] In step 520, the actual proportion of the overrepresented sequence is determined. This may be done as described above by taking the number of wells showing only that sequence and dividing by the number of informative wells. In step 530, the average concentration of at least one of the sequences is measured (the reference sequence). In one embodiment, the reference sequence is the overrepresented sequence. In another

embodiment, the reference sequence is the underrepresented sequence. The measurement may be done by counting the number of wells negative for the reference sequence in the digital PCR experiment. The relationship between the proportion of negative wells and the average template concentration is described by the Poisson distribution as described in the next subsection.

[0107] In step 540, an expected amount of wells positive for the different sequences is calculated, for example, using a Poisson distribution. The expected amount may be as a probability of the sequence per well, an average sequence per well, the number of wells containing the sequence or any other suitable quantity. In step 550, the expected P_r is calculated from the expected amounts. In step 560, a cutoff value is calculated from the expected P_r, for example, by using SPRT. In step 570, a classification of the nucleic acid sequence imbalance is determined. Specific aspects of method 500 are now described.

1. Determining expected amount of sequences

[0108] Once the average concentration per well (reaction or reaction mixture) is known from step 530, the expected number of wells showing that sequence may be calculated in step 540. This amount may be expressed as a %, a fractional value, or an integer value. Using a specific example for illustration, assume the average concentration of the reference template per well (m_r) is 0.5 per well and the genotype of the trisomy 21 fetus at the *PLAC4* SNP, rs8130833, is AGG. Therefore, the reference template would be the A allele and the overrepresented template would be the G allele.

[0109] In one embodiment, a Poisson distribution is assumed for the distribution of the A allele among the reaction mixtures of the wells of the measurement procedure, such as digital PCR. In other embodiments, other distribution functions are used, such as binomial distribution.

[0110] The Poisson equation is: $P(n) = \frac{m^n e^{-m}}{n!}$ where, n=number of template molecules per well; P(n)=probability of n template molecules in a particular well; and m= average number of template molecules in one well in a particular digital PCR experiment.

[0111] Accordingly, the probability of any well not containing any molecule of the A allele at an average A-allele concentration of 0.5 would be:

$$P(0) = \frac{0.5^0 e^{-0.5}}{0!} = e^{-0.5} = 0.6065.$$

[0112] Hence, the probability of any well containing at least one molecule of the A allele would be: $1 - 0.6065 = 0.3935$. Therefore, ~39% of the wells would be expected to contain at least one molecule of the A allele.

[0113] As for the non-reference nucleic acid sequence, for each cell of a trisomy 21 fetus, the genomic ratio of A to G would be 1:2. Assuming that the A to G ratio in the extracted RNA or DNA sample would remain unchanged, the average concentration of the G allele per well would be two times that of the A allele, i.e. $2 \times 0.5 = 1$.

[0114] Accordingly, the probability of any well not containing any molecule of the G allele at an average G-allele concentration of 1 would be:

$$P(0) = \frac{1^0 e^{-1}}{0!} = e^{-1} = 0.3679$$

[0115] Hence, the probability of any well containing at least one molecule of the G allele would be: $1 - 0.3679 = 0.6321$. Therefore, ~63% of the wells would be expected to contain at least one molecule of the G allele.

2. Determining proportion of overrepresented sequence

[0116] After the expected amounts are calculated, the proportion of the overrepresented nucleic acid sequence may be determined. Assuming that the filling of the wells with the A allele and the G allele are independent, the probability of a well containing both alleles would be $0.3935 \times 0.6321 = 0.2487$. Therefore, ~25% of the wells would be expected to contain both alleles.

[0117] The proportion of wells expected to contain the A allele but not the G allele would be the number of wells containing at least one A allele deducted by the number of wells containing both the A and G alleles: $0.3935 - 0.2487 = 0.1448$. Similarly, the proportion of wells expected to contain the G allele but not the A allele would be: $0.6321 - 0.2487 = 0.3834$. An informative well is defined as a well being positive for either the A allele or the G allele but not both.

[0118] Hence, the expected ratio of wells containing the A allele relative to the G allele in a digital RNA-SNP analysis is 0.1448/0.3834. In other words, the proportion of wells positive

only for the G allele is 2.65 times that of wells positive only for the A allele. This is in contrast to the fetal genomic ratio where the overrepresented allele is 2 times that of the other allele.

[0119] For SPRT analysis, the proportion of the informative wells positive for the overrepresented allele (P_r) is calculated and interpreted using SPRT curves. In the current example, the proportion of informative wells would be: 0.1448 + 0.3834 = 0.5282. Hence, the expected Pr of a trisomy 21 case at m_r 0.5 is: 0.3834/0.5282 = 0.73.

[0120] As the average template concentration (*m*) is a key parameter in the Poisson equation, the P_r would vary with *m*. FIG. 6 shows a table 600 that tabulates the expected digital RNA-SNP allelic ratio and P_r of trisomy 21 samples for a range of template concentrations expressed as the average reference template concentration per well (m_r) according to an embodiment of the present invention. Table 600 shows the expected allelic ratio and proportion of informative wells positive for the overrepresented allele (P_r) for a series of average reference template concentrations per well (m_r).

[0121] The expected value of P_r varies with the average concentration of the reference allele per well (m_r) in a non-linear fashion. As shown in table 600, the expected value of P_r for accepting the alternative hypothesis would increase with m_r. As the expected value of P_r for accepting the null hypothesis is fixed at 0.5, the samples with and without allelic imbalance would separate more widely in terms of the value of P_r when m_r increases. Note that in other embodiments the value for accepting the null hypothesis may be other than 0.5. This might occur when the normal ratio is different than 1:1, e.g., 5:3, and thus an imbalance would occur when the ratio deviates from 5:3. The difference in the amounts of the two different nucleic acid sequences would then be determined on a case by case basis.

[0122] However, as previous methods (e.g., Zhou, W. *et al.* 2002, supra) used a fixed expected value of P_r for LOH samples, they underestimated the value of P_r for those samples with LOH (alternative hypothesis accepted). The degree of underestimation would increase with m_r. In other words, the higher the average concentration of the reference allele in the DNA sample, the more inaccurate the old methods would be. This underestimation of P_r for accepting the alternative hypothesis would lead to the inaccurate calculation of the cutoff values for accepting both the null and alternative hypotheses.

3. Calculating cutoff values based on expected P_r

[0123] For embodiments using SPRT, one may use the equations for calculating the upper and lower boundaries of the SPRT curves from El Karoui at al. (2006). Furthermore, the level of statistical confidence preferred for accepting the null or alternative hypothesis could be varied through adjusting the threshold likelihood ratio in the equations. In this study, a threshold likelihood ratio of 8 is used because this value had been shown to provide satisfactory performance to discriminate samples with and without allelic imbalance in the context of cancer detection. Thus, in one embodiment, the equations for calculating the upper and lower boundaries of the SPRT curves are:

Upper boundary = $[(\ln 8)/N - \ln \delta]/\ln \gamma$

Lower boundary = $[(\ln 1/8)/N - \ln \delta]/\ln \gamma$

where, $\delta = (1 - \theta_1)/(1 - \theta_0)$

$\gamma = -(\theta_1(1 - \theta_0) / \theta_0(1 - \theta_1)$

θ_0 = proportion of informative wells containing the non-reference allele if the null hypothesis is true
= 0.5 (see below)

θ_1 = proportion of informative wells containing the non-reference (i.e. overrepresented) allele if the alternative hypothesis is true

N = number of informative wells
= number of wells positive for either allele only

(ln is a mathematical symbol representing the natural logarithm, i.e. $\log_e$.)

[0124] For the determination of θ_0 for accepting the null hypothesis, the sample is assumed to be obtained from a pregnant woman carrying a euploid fetus. Under this assumption, the expected number of wells positive for either template would be 1:1, and thus the expected proportion of informative wells containing the non-reference allele would be 0.5.

[0125] For the determination of θ_1 for accepting the alternative hypothesis, the sample is assumed to be obtained from a pregnant woman carrying a trisomy 21 fetus. The calculations for the expected P_r of trisomy 21 cases for digital RNA-SNP analysis are detailed in Table 600. Hence, θ_1 for digital RNA-SNP analysis refers to the data shown in the last column of table 600.

4. Measurement of average concentration

[0126] The measurement of m_r may be performed through a variety of mechanisms as known or will be known to one skilled in the art. In one embodiment, the value of m_r is determined during the experimental process of digital PCR analysis. As the relationship between the value of m_r and the total number of wells being positive for the reference allele can be governed by a distribution (e.g. the Poisson distribution), m_r can be calculated from the number of wells being positive for the reference allele using this formula:

$m_r = -\ln(1 - \text{proportion of wells being positive for the reference allele})$

Note that ln is the natural logarithm, i.e., $\log_e$. This approach provides a direct and precise estimation of m_r in the DNA sample used for the digital PCR experiment.

[0127] This method may be used to achieve a desired concentration. For example, the extracted nucleic acids of a sample may be diluted to a specific concentration, such as one template molecule per reaction well, as done in step 240 of method 200. In an embodiment using the Poisson distribution, the expected proportion of wells with no template may be calculated as e^{-m}, where m is the average concentration of template molecules per well. For example, at an average concentration of one template molecule per well, the expected proportion of wells with no template molecule is given by e^{-1}, i.e., 0.37 (37%). The remaining 63% of wells will contain one or more template molecules. Typically, the number of positive and informative wells in a digital PCR run would then be counted. The definition of informative wells and the manner by which the digital PCR data are interpreted depends on the application.

[0128] In other embodiments, the average concentration per well, m_r, is measured by another quantification method, for example, quantitative real-time PCR, semi-quantitative competitive PCR, real-competitive PCR using mass spectrometric methods, etc.

B. Digital RCD

[0129] Digital RCD using the average concentration may be performed in a similar manner to the digital SNP method described above. The numbers of wells positive for the reference chromosome (non-chromosome 21) marker, the chromosome 21 marker and both markers can be determined by digital PCR. The average concentration of the reference marker per well (m_r) can be calculated from the total number of wells negative for the reference marker,

irrespective of the positivity of the chromosome 21 marker, according to the Poisson probability function as in the calculation of m_r for the digital SNP analysis.

[0130] SPRT analysis may then be used for classifying a plasma sample as being obtained from a pregnant woman carrying a euploid or a trisomy 21 fetus. The null hypothesis would be accepted when the fetus was euploid. In this scenario, the expected ratio for the wells positive for the reference marker and chromosome 21 marker would be 1:1 and, thus, the expected proportion of informative wells with positive signal for chromosome 21 marker would be 0.5. The alternative hypothesis would be accepted when the fetus was trisomic for chromosome 21. In this scenario, if the sample DNA was solely derived from the fetus, the average concentration of the chromosome 21 marker in each well would be 3/2 times the average concentration of the reference marker (m_r).

[0131] While digital RCD may be used to determine chromosome dosage through the detection of fetal-specific markers, e.g. epigenetic signatures of the placenta (Chim, SSC. *et al.* 2005 *Proc Natl Acad Sci U S A* 102, 14753-14758), an embodiment of the digital RCD analysis uses non-fetal-specific markers. Thus, an additional step of measuring the percentage of fetal material would occur when non-fetal specific markers are used. Therefore, the average concentration of the chromosome 21 marker per well would be dependent on the proportion of the fetal DNA in the sample and can be calculated using: m_r [(200% + fetal DNA percentage)/200%].

[0132] Again using a specific example for illustration, the average concentration of the reference template, chromosome 1, per well (m_r) is assumed to be 0.5 and 50% of the DNA is assumed to be derived from the fetus and 50% of the DNA in the sample is derived from the mother.

[0133] Accordingly, using the Poisson distribution, the probability of any well not containing any molecule of the chromosome 1 locus when its average concentration is 0.5 per well would be:

$$P(0) = \frac{0.5^0 e^{-0.5}}{0!} = e^{-0.5} = 0.6065$$

[0134] Hence, the probability of any well containing at least one molecule of the chromosome 1 locus would be: $1 - 0.6065 = 0.3935$. Therefore, ~39% of the wells would be expected to contain at least one molecule of the locus.

[0135] For each cell of this trisomy 21 fetus, the genomic ratio of chromosome 21 to chromosome 1 would be 3:2. The ratio between chromosome 21 and chromosome 1 in the DNA sample would be dependent on the fractional fetal DNA concentration (fetal DNA %) and would be: 3 × fetal DNA % + 2 (1 – fetal DNA %) : 2 × fetal DNA % + 2 × (1 – fetal DNA %). Thus, in this case when the fractional fetal DNA concentration is 50%, the ratio would be: (3 × 50% + 2 × 50%) / (2 × 50% + 2 × 50%) = 1.25. If the digital SNP method did not use fetal specific markers, such a calculation could also be used to calculate the average concentration of the non-reference sequence.

[0136] Hence, when the average concentration of the chromosome 1 locus per well is 0.5, the average concentration of the chromosome 21 locus per well is: 1.25 × 0.5 = 0.625. Accordingly, the probability of any well not containing any molecule of the chromosome 21 locus when its average concentration is 0.625 per well would be:

$$P(0) = \frac{0.625^0 e^{-0.625}}{0!} = e^{-0.625} = 0.5353$$

[0137] Hence, the probability of any well containing at least one molecule of the chromosome 21 locus would be: 1 – 0.5353 = 0.4647. Therefore, ~46% of the wells would be expected to contain at least one molecule of the locus. Assuming that the filling of the wells with either loci are independent, the probability of a well containing both loci would be 0.3935 × 0.4647 = 0.1829. Therefore, ~18% of the wells would be expected to contain both loci.

[0138] The proportion of wells expected to contain the chromosome 1 locus but not the chromosome 21 locus would be the number of wells containing at least one chromosome 1 locus deducted by the number of wells containing both loci: 0.3935 – 0.1829 = 0.2106. Similarly, the proportion of wells expected to contain the chromosome 21 locus but not both loci would be: 0.4647 – 0.1829 = 0.2818. An informative well is defined as a well positive for either the chromosome 1 locus or the chromosome 21 locus but not both.

[0139] Hence, the expected chromosome 21 to chromosome 1 ratio in digital RCD analysis is 0.2818/0.2106 = 1.34. In other words, the proportion of wells positive only for the chromosome 21 locus is 1.34 times that of wells positive only for the chromosome 1 locus. This is in contrast to the ratio of 1.25 in the DNA sample.

[0140] For SPRT analysis, the proportion of the informative wells positive for the chromosome 21 locus (P_r) would need to be calculated and interpreted using SPRT curves. In the current example, the proportion of informative wells would be: 0.2106 + 0.2818 = 0.4924. Hence, the expected Pr of a trisomy 21 case with 50% fetal DNA at m_r 0.5 is: 0.2818/0.4924 = 0.57.

[0141] As the average template concentration (m) is a key parameter in the Poisson equation, the P_r would vary with m. FIG. 7 shows a table 700 that tabulates the expected P_r for the fractional fetal DNA concentrations of 10%, 25%, 50% and 100% in trisomy 21 samples at a range of template concentrations expressed as the average reference template concentration per well (m_r) according to an embodiment of the present invention. The calculations for the expected P_r of trisomy 21 cases for digital RCD analyses are detailed in table 700. Hence, θ_1 for digital RCD analysis of samples with varying fetal DNA fractional concentrations can be obtained from the columns showing the corresponding expected P_r-values in table 700.

C. Results

1. Comparison of different m_r

[0142] The basis for the difference between the theoretical (as in the fetal genome) and experimentally-expected degree of allelic or chromosomal imbalance and the calculations to determine the latter for a range of m_r values are shown in tables 600 and 700. In digital RNA-SNP analysis of a trisomy 21 sample, when m_r=0.5, wells containing just the overrepresented allele with respect to wells containing just the reference allele, namely the digital RNA-SNP ratio, is 2.65 (table 600). In digital RCD analysis of a specimen composed of 100% fetal DNA, when m_r=0.5, wells positive solely for the chromosome 21 locus with respect to those positive solely for the chromosome 1 locus, namely the digital RCD ratio, is 1.7 (table 700) (P_r = 0.63, thus the digital RCD ratio is 0.63/(1-0.63) = 1.7). As the fractional fetal DNA concentration decreases, the digital RCD ratio decreases for the same m_r (table 700).

[0143] As shown in tables 600 and 700, the extent of allelic or chromosomal overrepresentation increases with m_r. However, the percentage of informative wells approaches its maximum around m_r=0.5 and decreases gradually with further increase in m_r. In practice, the decline in the proportion of informative wells could be compensated by increasing the total number of wells analyzed if the amount of template molecule for the

specimen is not limiting, but additional wells would require an increase in reagent costs. Hence, optimal digital PCR performance is a trade-off between the template concentration and total number of wells tested per sample.

2. Example using SPRT curves

[0144] As discussed above, the expected degree of allelic or chromosomal imbalance for a digital PCR experiment is dependent on the actual template concentration per reaction mixture (e.g. a well). We describe the template concentration based on the reference allele, i.e. the average reference template concentration per well (m_r). As shown in the above equation, the expected P_r can be used to determine the plotting of the upper and the lower SPRT curves. Since the expected P_r is in turn dependent on the value of m_r, the plotting of the SPRT curves would essentially be dependent on the value of m_r. Thus, in practice, a set of SPRT curves relevant for the actual m_r of a digital PCR dataset would need to be used for the interpretation of the P_r from that particular run.

[0145] FIG. 8 shows a plot 800 illustrating the degree of differences in the SPRT curves for m_r values of 0.1, 0.5 and 1.0 for digital RNA-SNP analysis according to an embodiment of the present invention. Each set of digital PCR data should be interpreted with the specific curves relevant to the exact m_r value of that particular run. Note that since the expected degree of allelic or chromosomal imbalance for the digital RNA-SNP and RCD approaches are different (2:1 for the former and 3:2 for the latter), different sets of SPRT curves are needed for the two digital PCR systems. The experimentally derived P_r is interpreted with the relevant SPRT curves selected by the corresponding m_r of the digital PCR run. This is in contrast to the previous reported use of SPRT for molecular detection of LOH by digital PCR where a fixed set of curves had been used.

[0146] The practical manner for interpreting the digital PCR data using SPRT is illustrated below using a hypothetical digital RNA-SNP run. After digital RNA-SNP analysis of each case, the number of wells positive for the A allele only, the G allele only or both alleles are counted. The reference allele is defined as the allele with the smaller number of positive wells. The value of m_r is calculated using the total number of wells negative for the reference allele, irrespective of whether the other allele is positive, according to the Poisson probability density function. The data of our hypothetical example are as follows:

[0147] In a 96-well reaction, 20 wells are positive for the A allele only, 24 wells are positive for the G allele only, and 33 wells are positive for both alleles. The A allele is regarded as the reference allele because there are less A-positive than G-positive wells. The number of wells negative for the reference allele is 96 – 20 – 33 = 43. Therefore, m_r can be calculated using the Poisson equation: $-\ln(43/96) = 0.80$. The experimentally determined P_r of this case is: 24 / (20 + 24) = 0.55.

[0148] According to table 600, the expected P_r of a trisomy 21 sample at $m_r = 0.8$ is 0.76. Thus, θ_1 is 0.76 for this case. The SPRT curves based on $\theta_1 = 0.76$ would be used to interpret the experimentally derived P_r of this case which is 0.55. When P_r=0.55 is fitted onto the relevant SPRT curves, the data point falls under the lower curve. Hence, the case is classified as euploid, see FIG. 3.

3. Comparison to old method

[0149] FIG. 9A shows a table 900 of a comparison of the effectiveness of the new and old SPRT algorithms for classifying euploid and trisomy 21 cases in 96-well digital RNA-SNP analyses. FIG. 9B shows a table 950 of a comparison of the effectiveness of the new and old SPRT algorithms for classifying euploid and trisomy 21 cases in 384-well digital RNA-SNP analyses. The new algorithm refers to the selection of SPRT curves specific for the m_r derived from the digital PCR data. The old algorithm refers to the use of a fixed set of SPRT curves for all digital PCR runs. The effect of incorrect calculation of the cutoff values on the accuracy of classification is revealed by the simulation analysis shown in table 900.

[0150] Compared with the use of a fixed set of SPRT curves in previous studies, the proportion of unclassifiable data is much lower with our approach, as shown in tables 900 and 950. For example, using our approach, at m_r=0.5, 14% and 0% of trisomy 21 samples would be unclassifiable for 96-well and 384-well digital RNA-SNP analyses, respectively, but 62% and 10%, respectively, with the use of fixed curves (Table 900). Hence, our approach allows disease classification with lesser number of informative wells.

[0151] As shown in table 900, the new algorithm is more accurate in classifying the samples as having or not having allelic ratio skewing for all values of m_r from 0.1 to 2.0. For example, when m_r equals 1.0 and a 96–well digital RNA-SNP run is performed, the new algorithm correctly classifies 88% and 92% of samples with and without allelic ratio

skewing, respectively, whereas the percentage of correct classification of samples with and without allelic ratio skewing using the old algorithm is only 19% and 36%, respectively.

[0152] Using the new algorithm, the separation of samples with and without allelic ratio skewing would increase with m_r. As a result, the classification accuracies would increase with m_r. The effect of increase in separation of the two groups of samples on the classification accuracy would diminish when m_r increases to beyond 2.0 because of the reduction in the percentage of informative wells. In contrast, using the old algorithm, the classification accuracies significantly reduce when m_r increases because of the increased deviation of expected P value from its true value.

[0153] Our experimental and simulation data show that digital RNA-SNP is an effective and accurate method for trisomy 21 detection. As *PLAC4* mRNA in maternal plasma is derived purely from the fetus, for 12 of the 13 maternal plasma samples tested, only one 384-well digital PCR experiment was required for correct classification (Table 1350 of FIG. 13B). This homogenous, real-time digital PCR-based approach thus offers an alternative to the mass spectrometry-based approach for RNA-SNP analysis (Lo, YMD, *et al.* 2007 *Nat Med*, supra). Apart from placental-specific mRNA transcripts, we also envision that other types of fetal-specific nucleic acid species in maternal plasma could be used for digital PCR-based detection of fetal chromosomal aneuploidies. One example is fetal epigenetic markers (Chim, SSC et al. (2005) *Proc Natl Acad Sci U S A* 102, 14753-14758; Chan, KCA *et al.* (2006) *Clin Chem* 52, 2211-2218), which have recently been used for the noninvasive prenatal detection of trisomy 18 using the epigenetic allelic ratio (EAR) approach (Tong, YK et al. (2006) *Clin Chem* 52, 2194-2202). Thus, we predict that digital EAR would be a possible analytical technique.

V. INCREASING %, MULTIPLE MARKERS, AND PCR ALTERNATIVES

[0154] As described above, the application of embodiments of the present invention to DNA extracted from maternal plasma can be complicated when the fetal DNA only constitutes a minor fraction of maternal plasma DNA, with a mean fractional concentration of some 3% between weeks 11 and 17 of gestation. Nevertheless, as shown herein, digital RCD allows aneuploidy detection even when the aneuploid DNA is present as a minor population. With a decreasing fractional concentration of fetal DNA, such as may be present during early gestation, a larger number of informative counts is needed for digital RCD. A significance of the present work, as summarized in table 1200 of FIG. 12, is that we have provided a set of

benchmark parameters, e.g. fractional fetal DNA and total template molecules required, which diagnostic assays can be built upon. In our opinion, 7680 total number of reactions for a fractional fetal DNA concentration of 25% is a particularly attractive set of benchmark parameters. These parameters should allow euploid and trisomy 21 samples to be classifiable correctly 97% of the time, as shown in table 1200.

[0155] The number of plasma DNA molecules that are present per unit volume of maternal plasma is limited (Lo, YMD. *et al.* 1998 *Am J Hum Genet* 62, 768-7758). For example, in early pregnancy, the median maternal plasma concentration of an autosomal locus, the *β-globin* gene, has been shown to be 986 copies/mL, with contributions from both the fetus and mother (Lo, YMD. *et al.* 1998 *Am J Hum Genet* 62, 768-7758). To capture 7,680 molecules, DNA extracted from some 8 mL of maternal plasma would be needed. This volume of plasma, obtainable from ~15 mL of maternal blood, is at the limit of routine practice. However, we envision that multiple sets of chr21 and reference chromosome targets can be combined for digital RCD analysis. For 5 pairs of chr21 and reference chromosome targets, just 1.6 mL of maternal plasma would be needed to provide the number of template molecules needed for analysis. Multiplex single molecule PCR could be performed. The robustness of such multiplex single molecule analysis has been demonstrated previously for single molecule haplotyping (Ding, C. and Cantor, CR. 2003 *Proc Natl Acad Sci U S A* 100, 7449-7453).

[0156] Alternatively, to achieve a fractional fetal DNA concentration of 25%, methods may allow the selective enrichment of fetal DNA (Li, Y. et al. 2004 *Clin Chem* 50, 1002-1011) or the suppression of the maternal DNA background (Dhallan, R et al. 2004 *JAMA* 291, 1114-1119) or both, in maternal plasma. Apart from such physical methods for fetal DNA enrichment and maternal DNA suppression, it would also be possible to use a molecular enrichment strategy, such as by targeting fetal DNA molecules which exhibit a particular DNA methylation pattern (Chim, SSC et al, 2005 *Proc Natl Acad Sci USA* 102,14753-14758, Chan, KCA et al. 2006 *Clin Chem* 52, 2211-2218; Chiu, RWK *et al.* 2007 *Am J Pathol* 170, 941-950.)

[0157] Additionally, there are now a number of alternative approaches to the manual set up of digital real-time PCR analyses as used in the current study for conducting digital PCR. These alternative approaches include microfluidics digital PCR chips (Warren, L et al. 2006 *Proc Natl Acad Sci U S A* 103, 17807-17812; Ottesen, EA et al. 2006 *Science* 314, 1464-

1467), emulsion PCR (Dressman, D et al. 2003 *Proc Natl Acad Sci U S A* 100, 8817-8822), and massively parallel genomic sequencing (Margulies, M. *et al.* 2005 *Nature* 437, 376-380) using for example the Roche 454 platform, the Illumina Solexa platform, and the SOLiD™ system of Applied Biosystems, etc. With regard to the latter, our method is also applicable to massively parallel sequencing methods on single DNA molecules, which do not require an amplification step, such as the Helicos True Single Molecule DNA sequencing technology (Harris TD et al. 2008 Science, 320, 106-109), the single molecule, real-time (SMRT™) technology of Pacific Biosciences, and nanopore sequencing (Soni GV and Meller A. 2007 *Clin Chem* 53, 1996-2001). With the use of these methods, digital RNA-SNP and digital RCD could be performed rapidly on a large number of samples, thus enhancing the clinical feasibility of the methods proposed here for noninvasive prenatal diagnosis.

EXAMPLES

The following examples are offered to illustrate, but not to limit the claimed invention.

I. COMPUTER SIMULATIONS

[0158] Computer simulation was performed to estimate the accuracy of diagnosing trisomy 21 using the SPRT approach. The computer simulation was performed with the Microsoft Excel 2003 software (Microsoft Corp., USA) and SAS 9.1 for Windows software (SAS Institute Inc., NC, USA). The performance of digital PCR is an interplay between the reference template concentration (m_r), number of informative counts and projected degree of allelic or chromosomal imbalance (P_r). Separate simulations were performed for a range of each of these variables. Since the decision boundaries of the SPRT curves for digital RNA-SNP and digital RCD were different, the simulation analyses for the two systems were performed separately.

[0159] For each digital PCR condition simulated (i.e. m_r, fetal DNA fractional concentration, total well number), two rounds of simulation were performed. The first round simulated the scenario that the tested samples were obtained from pregnant women carrying euploid fetuses. The second round simulated the scenario when the tested samples were obtained from pregnant women carrying trisomy 21 fetuses. For each round, testing of 5000 fetuses was simulated.

A. RNA-SNP

[0160] For digital RNA-SNP, simulations of a 384-well experiment with m_r=0.1 to m_r=2.0 were performed. At each m_r value, we simulated the scenario whereby 5000 euploid and 5000 trisomy 21 fetuses were tested. The SPRT curves appropriate for the given m_r were used to classify the 10,000 fetuses. FIG. 10 is a table 1000 showing the percentages of fetuses correctly and incorrectly classified as euploid or aneuploid and those not classifiable for the given informative counts according to an embodiment of the present invention. The accuracies for diagnosing euploid and aneuploid cases are both 100%, for m_r between 0.5 and 2.0. When m_r=0.1, only 57% and 88% of euploid and trisomy 21 fetuses could be accurately classified after the analysis of 384 wells.

[0161] The simulation data were generated as described in the following steps:

[0162] In step 1, for each well, two random numbers were generated using the Random(Poisson) function of the SAS program (www.sas.com/technologies/analytics/statistics/index.html) to represent the A and the G alleles, respectively. The Random(Poisson) function would generate positive integers starting from 0 (i.e. 0, 1, 2, 3,....) and the probability of each integer being generated was dependent on the probability of this number according to the Poisson probability density function for a given mean value which represented the average concentration of the alleles per well. A well was regarded as positive for the A allele if the random number representing the A allele was larger than zero, i.e. contained 1 or more molecules of the A allele. Similarly, the well was regarded as positive for the G allele if the random number representing the G allele was larger than zero.

[0163] To simulate the scenario of a pregnant woman carrying a euploid fetus, the same mean value was used for generating the random numbers for the A allele and the G allele. For example, in the analysis simulating digital RNA-SNP analyses at $m_r = 0.5$, the mean value for either the A allele of the G allele was set identically to 0.5 which meant an average concentration for either allele of 0.5 molecule per well. Using the Poisson equation, at a mean concentration of 0.5, the proportion of wells being positive for the A or the G alleles would be the same and was 0.3935, see table 600.

[0164] When simulating the digital RNA-SNP analysis of a pregnant woman with a trisomy 21 fetus at $m_r = 0.5$, the average concentration of the overrepresented allele per well

would be expected to be 2 times of that of the reference allele, i.e. 1. In this situation, the probability of a well being positive for the overrepresented allele was 0.6321, see table 600.

[0165] After generating a random number for a digital PCR well, the well could be classified as one of the following statuses:

a. negative for both the A and the G alleles
b. positive for both the A and the G alleles
c. positive for the A allele but negative for the G allele
d. positive for the G allele but negative for the A allele

[0166] In step 2, step 1 was repeated until the desired number of wells, 384 wells for the current simulation, had been generated. The numbers of wells positive for the A allele only and the G allele only were counted. The allele with less positive wells was regarded as the reference allele and the allele with more positive wells was regarded as the potentially overrepresented allele. The number of the informative wells was the total number of wells positive for either allele but not both. The proportion of informative wells containing the potentially overrepresented allele (P_r) was then calculated. The upper and lower boundaries for the relevant SPRT curves to accept the null or alternative hypothesis were calculated according to an embodiment of the present invention.

[0167] In step 3, 5000 simulations were performed for each of the two scenarios of the pregnant woman carrying a euploid or a trisomy 21 fetus. Each simulation could be regarded as an independent biological sample obtained from pregnant women. In Table 1000, the correct classification of euploid cases refers to those euploid cases in which the null hypothesis was accepted and the incorrect classification of euploid cases refers to those euploid cases in which the alternative hypothesis was accepted. Similarly, those trisomy 21 cases in which the alternative hypothesis was accepted were regarded as correctly classified and those trisomy 21 cases in which the null hypothesis was accepted were regarded as incorrectly classified. For both groups, those cases in which neither the null or alternative hypothesis was accepted after the pre-specified total number of wells had been simulated were regarded as unclassified.

[0168] In step 4, steps 1 to 3 were performed for m_r ranging from 0.1 to 2.0 at increments of 0.1.

B. RCD

[0169] FIG. 11 is a table 1100 showing computer simulations for digital RCD analysis for a pure (100%) fetal DNA sample for m_r ranging from 0.1 to 2.0 according to an embodiment of the present invention. As the fractional fetal DNA concentration becomes lower, the degree of chromosome 21 overrepresentation diminishes and thus a larger number of informative wells for accurate disease classification is required. Hence, simulations were further performed for fetal DNA concentrations of 50%, 25% and 10% for a total well number ranging from 384 to 7680 wells at m_r=0.5.

[0170] FIG. 12 is a table 1200 showing results of computer simulation of accuracies of digital RCD analysis at m_r=0.5 for the classification of samples from euploid or trisomy 21 fetuses with different fractional concentrations of fetal DNA according to an embodiment of the present invention. The performance of digital RCD is better for cases with a higher fetal DNA fractional concentration. At 25% fetal DNA concentration and a total number of 7680 PCR analyses, 97% of both euploid and aneuploid cases would be classifiable with no incorrect classification. The remaining 3% of cases require further analyses until classification can be achieved.

[0171] The procedures for simulating digital RCD analyses were similar to those described for digital RNA-SNP analysis. The steps for the simulations are described below:

[0172] In step 1, two random numbers under the Poisson probability density function were generated to represent the reference locus, chromosome 1, and the chromosome 21 locus. For subjects carrying euploid fetuses, the average concentrations of both the chromosome 1 and chromosome 21 loci were the same. In this simulation analysis, an average template concentration of 0.5 for each locus per well was used. For subjects carrying trisomy 21 fetuses, the m_r in this simulation was 0.5 but the average concentration of the chromosome 21 locus per well would depend on the fractional fetal DNA concentration in the tested sample, as shown in Table 700. The distribution of the reference and/or the chromosome 21 loci to a well was determined by the random numbers representing the respective locus which were generated according to the Poisson probability density function with the appropriate average concentration of the locus per well.

[0173] In step 2, step 1 was repeated until the desired number of wells had been generated, e.g. 384 wells for a 384-well plate experiment. The numbers of wells positive for chromosome 1 only and chromosome 21 only were counted. The number of the informative

wells was the total number of wells positive for either one of the chromosomes but not both. The proportion of informative wells positive for chromosome 21 (P_r) was then calculated. The upper and lower boundaries of the relevant SPRT curves to accept the null or alternative hypothesis were calculated as described in the earlier section on SPRT analysis.

[0174] In step 3, 5000 simulations were performed for each of the two scenarios of the pregnant woman carrying a euploid or a trisomy 21 fetus. Each simulation could be regarded as an independent biological sample obtained from pregnant women. In Table 1100, the correct classification of euploid cases refers to those euploid cases in which the null hypothesis was accepted and the incorrect classification of euploid cases refers to those euploid cases in which the alternative hypothesis was accepted. Similarly, those trisomy 21 cases in which the alternative hypothesis was accepted were regarded as correctly classified and those trisomy 21 cases in which the null hypothesis was accepted were regarded as incorrectly classified. For both groups, those cases in which neither the null or alternative hypothesis was accepted after the pre-specified total number of wells had been simulated were regarded as unclassified.

[0175] In step 4, steps 1 to 3 were repeated for samples with 10% 25%, 50% and 100% fetal DNA at total well numbers ranging from 384 to 7680.

II. VALIDATION OF TRISOMY 21 DETECTION

A. RNA-SNP for *PLAC4*

[0176] The practical feasibility of digital RNA-SNP was demonstrated using the rs8130833 SNP on the *PLAC4* gene on chromosome 21 (Lo, YMD et al. 2007 *Nat Med* 13, 218-223). Placental DNA and RNA samples from two euploid and two trisomy 21 heterozygous placentas were analyzed. The placental DNA samples were analyzed with the digital RNA-SNP protocol but with the omission of the reverse transcription step, thus essentially converting the procedure to digital *DNA*-SNP analysis. To strike a balance between the chance of correct case classification and the proportion of informative wells, we diluted the samples aiming for one allele of any type per well and confirmed by a 96-well digital PCR analysis. This was followed by a 384-well digital RNA-SNP experiment. P_r and m_r were calculated and the SPRT curve for this m_r value was used for data interpretation.

[0177] FIG. 13A shows a table 1300 of digital RNA-SNP analysis in placental tissues of euploid and trisomy 21 pregnancies according to an embodiment of the present invention.

Genotypes were determined by mass spectrometric assay. "Euploid" was assigned when the experimentally obtained P_r was below the unclassifiable region; "T21", representing trisomy 21, was assigned when the experimentally obtained P_r was above the unclassifiable region. T21, trisomy 21. Each of these cases was correctly classified, using both the DNA and RNA samples, with one 384-well experiment.

[0178] We further tested plasma RNA samples from nine women carrying euploid fetuses and four women carrying trisomy 21 fetuses. FIG. 13B shows a table 1350 of digital RNA-SNP analysis of maternal plasma from euploid and trisomy 21 pregnancies according to an embodiment of the present invention. All of the cases were correctly classified. Initial results for one trisomy 21 case (M2272P) fell within the unclassifiable region between the SPRT curves after one 384-well experiment. Thus, an additional 384-well run was performed. New m_r and P_r values were calculated from the aggregated data from the total of 768 wells and the classification was performed using a new set of SPRT curves selected based on this m_r value. The case was then scored correctly as aneuploid.

[0179] Our experimental and simulation data show that digital RNA-SNP is an effective and accurate method for trisomy 21 detection. As *PLAC4* mRNA in maternal plasma is derived purely from the fetus, for 12 of the 13 maternal plasma samples tested, only one 384-well digital PCR experiment was required for correct classification. This homogenous, real-time digital PCR-based approach thus offers an alternative to the mass spectrometry-based approach for RNA-SNP analysis. Apart from placental-specific mRNA transcripts, we also envision that other types of fetal-specific nucleic acid species in maternal plasma could be used for digital PCR-based detection of fetal chromosomal aneuploidies. One example is fetal epigenetic markers which have recently been used for the noninvasive prenatal detection of trisomy 18 using the epigenetic allelic ratio (EAR) approach (Tong YK et al. 2006 *Clin Chem,* 52, 2194-2202). Thus, we predict that digital EAR would be a possible analytical technique.

B. RCD

[0180] The practical feasibility of digital RCD for trisomy 21 detection was also investigated using a PCR assay targeting paralogous sequences on chromosome 21 and 1. Paralogous loci were used here by way of examples. Non-paralogous sequences on chromosome 21 and any other reference chromosome can also be used for RCD. Placental DNA samples from two euploid and two trisomy 21 placentas were diluted to approximately

one target template from either chromosome per well and confirmed by a 96-well digital PCR analysis. Each confirmed sample was analyzed by a 384-well digital RCD experiment and the values of P_r and m_r were calculated. For digital RCD, the chromosome 1 paralog was the reference template. This m_r value was used to select a corresponding set of SPRT curves for data interpretation. All of the placental samples were correctly classified as shown in FIG. 14A.

[0181] To demonstrate that the digital RCD approach can be used to detect trisomy 21 DNA which is mixed with an excess of euploid DNA, such as the scenario of fetal DNA in maternal plasma, mixtures containing 50% and 25% of trisomy 21 placental DNA in a background of euploid maternal blood cell DNA were analyzed. Placental DNA from 10 trisomy 21 and 10 euploid cases was each mixed with an equal amount of euploid maternal blood cell DNA, thus producing twenty 50% DNA mixtures. FIG. 14B shows a plot 1440 illustrating the SPRT interpretation for RCD analysis of the 50% fetal DNA mixtures according to an embodiment of the present invention. Similarly, placental DNA from 5 trisomy 21 and 5 euploid cases was each mixed with 3 times excess of euploid maternal blood cell DNA, thus producing ten 25% DNA mixtures. FIG. 14C shows a plot 1470 illustrating the SPRT interpretation for RCD analysis of the 25% fetal DNA mixtures. All the euploid and aneuploid DNA mixtures were correctly classified, as shown in FIGS. 14B and 14C.

[0182] Each sample reached the point of being classifiable after a number of 384-well digital PCR analyses as marked on FIGS. 14B and 14C. For the 50% DNA mixtures, the number of 384-well plates required ranged from one to five. For the 25% DNA mixtures, the number of 384-well plates required ranged form one to seven. The cumulative proportion of cases being correctly classified with the addition of every 384 digital PCR analyses were as predicted by the computer simulation presented in Table 1200.

III. METHOD WITH DIGITAL PCR

A. Digital RNA-SNP

[0183] All RNA samples were first reverse transcribed with a gene-specific reverse transcription primer using the ThermoScript reverse transcriptase (Invitrogen). Sequence of the reverse transcription primer was 5′-AGTATATAGAACCATGTTTAGGCCAGA-3′ (Integrated DNA Technologies, Coralville, IA). The subsequent treatment of the reverse transcribed RNA (i.e. the cDNA) samples for digital RNA-SNP, and DNA samples (e.g.

placental DNA) was essentially the same. Prior to digital PCR analysis, DNA and the cDNA samples were first quantified using a real-time PCR assay towards *PLAC4*, consisting of primers 5′-CCGCTAGGGTGTCTTTTAAGC-3′, 5′-GTGTTGCAATACAAAATGAGTTTCT-3′, and the fluorescent probe 5′-(FAM)ATTGGAGCAAATTC(MGBNFQ)-3′ (Applied Biosystems, Foster City, CA), where FAM is 6-carboxyfluorescein and MGBNFQ is a minor groove binding non-fluorescent quencher.

[0184] A calibration curve was prepared by serial dilutions of HPLC-purified single-stranded synthetic DNA oligonucleotides (Proligo, Singapore) specifying the amplicon. The sequence was 5′-CGCCGCTAGGGTGTCTTTTAAGCTATTGGAGCAAATTCAAATTTGGCTTAAAGAAAAAGAAACTCATTTTGTATTGCAACACCAGGAGTATCCCAAGGGACTCG-3′. The reaction was set up using 2X TaqMan Universal PCR Master Mix (Applied Biosystems) in a reaction volume of 25 μL. 400 nM of each primer and 80 nM of the probe were used in each reaction. The reaction was initiated at 50°C for 2 min, followed by 95°C for 10 min and 45 cycles of 95°C for 15s and 60°C for 1 min in an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Serial dilutions of the DNA or cDNA samples were then undertaken such that the subsequent digital PCR amplification could be performed at approximately one template molecule per well. At such a concentration, it was expected that approximately 37% of the reaction wells would show negative amplification and was first confirmed by conducting a 96-well digital real-time PCR analysis. This was followed by digital RNA-SNP analysis conducted in 384-well plates using a set of non-intron spanning primers: the forward primer 5′-TTTGTATTGCAACACCATTTGG-3′ and the gene-specific reverse transcription primer described above.

[0185] Two allele-specific TaqMan probes targeting each of the two alleles of the rs8130833 SNP on the *PLAC4* sequence were designed. Their sequences were 5′-(FAM)TCGTCGTCTAACTTG(MGBNFQ)-3′ and 5′-(VIC)ATTCGTCATCTAACTTG(MGBNFQ) for the G and A alleles, respectively. The reaction was set up using 2X TaqMan Universal PCR Master Mix in a reaction volume of 5 μL. Each reaction contains 1X TaqMan Universal PCR Master Mix, 572 nM of each primer, 107 nM of the allele-G-specific probe and 357 nM of the allele-A-specific probe. The reaction was carried out in the ABI PRISM 7900HT Sequence Detection System. The reaction was initiated at 50°C for 2 min, followed by 95°C for 10 min and 45 cycles of 95°C

for 15s and 57°C for 1 min. During the reaction, the fluorescence data were collected by the "Absolute Quantification" application of the SDS 2.2.2 software (Applied Biosystems). The software automatically calculated the baselines and the threshold values. The number of wells which were positive for either the A or the G alleles was recorded and subjected to SPRT analysis.

B. Digital RCD analysis

[0186] All placental and maternal buffy coat DNA samples used in this study were first quantified by the NanoDrop spectrophotometer (NanoDrop Technology, Wilmington, DE). The DNA concentration is converted to copies/μL using a conversion of 6.6 pg/cell. The amount of DNA corresponding to approximately one template per well was determined by serially diluting the DNA samples and confirmed with the real-time PCR assay in a 96-well format where we expect approximately 37% of the wells to show negative amplification. The PCR setup for the confirmatory plate was the same as described below except that only the probe for the reference chromosome was added. In the digital RCD analysis, the paralogous loci on chromosome 21 and 1 (Deutsch, S. *et al.* 2004 *J Med Genet* 41, 908-915) were first co-amplified by forward primer 5'-GTTGTTCTGCAAAAAACCTTCGA-3' and reverse primer 5'-CTTGGCCAGAAATACTTCATTACCATAT-3'. Two chromosome-specific TaqMan probes were designed to target the chromosome 21 and 1 paralogs, and their sequences were 5'-(FAM)TACCTCCAT AATGAGTAAA(MGBNFQ)-3' and 5'-(VIC)CGTACCTCTGTAATGTGTAA(MGBNFQ)-3', respectively. Each reaction contained 1X TaqMan Universal PCR Master Mix (Applied Biosystems), 450 nM of each primer, and 125 nM of each probe. The total reaction volume was 5μL/well. The reaction was initiated at 50°C for 2 min, followed by 95°C for 10 min and 50 cycles of 95°C for 15s and 60°C for 1 min. All real-time PCR experiments were carried out on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems), and the fluorescence data were collected by the "Absolute Quantification" application of the SDS 2.2.2 software (Applied Biosystems). The default baselines and manual threshold values were used. The number of wells which were positive for either chromosome 21 or chromosome 1 was recorded and subjected to SPRT analysis. One or more 384-well plates would be analyzed until disease classification was possible by SPRT.

IV. USING MICROFLUIDICS-BASED DIGITAL PCR

A. Digital RNA-SNP

[0187] This example demonstrates the performance of digital PCR analysis using microfluidics-based digital PCR. A variant of this approach is illustrated here, by way of example but not by way of limitation to, using a Fluidigm BioMark™ System. This system can perform over 9000 digital PCRs per run.

[0188] Placental tissues and maternal peripheral blood samples were obtained from pregnant women carrying euploid or trisomy 21 fetuses. Genotyping of the rs8130833 SNP on the *PLAC4* gene was carried out in placental DNA samples by primer extension followed by mass spectrometry. RNA was extracted from the placental and maternal plasma samples.

[0189] All RNA samples were reverse transcribed with a gene-specific reverse transcription primer (5′-AGTATATAGAACCATGTTTAGGCCAGA-3′) using the ThermoScript reverse transcriptase (Invitrogen). For the placental cDNA samples, serial dilutions were performed such that the subsequent digital PCR amplification could be performed at approximately one template molecule per well.

[0190] Digital PCR was conducted on the BioMark System™ (Fluidigm) with a 12.765 Digital Array (Fluidigm). Each Digital Array consists of 12 panels for accommodating 12 sample-assay mixtures. Each panel is further partitioned into 765 wells for carrying out a 7-nL reaction/well. The rs8130833 SNP region on the *PLAC4* gene was amplified by the forward primer (5′-TTTGTATTGCAACACCATTTGG-3′) and the gene-specific reverse transcription primer described above. Two allele-specific TaqMan probes targeting each of the two alleles of the rs8130833 SNP were designed. Their sequences were 5′-(FAM)TCGTCGTCTAACTTG(MGBNFQ)-3′ and 5′-(VIC)ATTCGTCATCTAACTTG(MGBNFQ) for the G and A alleles, respectively. The reaction for one array panel was set up using 2X TaqMan Universal PCR Master Mix in a reaction volume of 10 μL. Each reaction contains 1X TaqMan Universal PCR Master Mix, 572 nM of each primer, 53.5 nM of the allele-G-specific probe, 178.5 nM of the allele-A-specific probe and 3.5 μL of the cDNA sample. One reaction panel was used for each placental cDNA sample while 12 panels were used for each maternal plasma sample. The sample-assay mixtures were loaded into the Digital Array by a NanoFlex™ IFC controller (Fluidigm). The reaction was carried out in the BioMark™ System. The reaction was

initiated at 50°C for 2 min, followed by 95°C for 10 min and 40 cycles of 95°C for 15s and 57°C for 1 min.

[0191] Placental RNA samples from one euploid and two T21 heterozygous placentas were analyzed in a 765-well reaction panel. For each sample, the number of informative wells, comprising the ones positive for either the A or the G allele (but not both), was counted. The proportion of the overrepresented allele among all the informative wells (Pr) was determined. SPRT curves appropriate for the exact average reference template concentration per well (m_r) of these runs were applied to determine if the experimentally-obtained Pr indicated a euploid or T21 sample. As shown in FIG. 15A, all RNA samples were correctly classified using this approach.

[0192] We further tested the plasma RNA samples from four women carrying euploid fetuses and one woman carrying a trisomy 21 fetus. Each sample was analyzed in twelve 765-well reaction panels, i.e. 9180 reactions per plasma RNA sample. FIG. 15B shows the number of informative wells for each of the 12 panels for this plasma RNA sample. As shown in the table, the template concentration in the plasma sample was so diluted that the number of informative wells in any one reaction panel was not sufficient for the SPRT classification. The informative wells from three reaction panels had to be combined before this sample was classified as a euploid sample (FIG. 15C). FIG. 15C shows that using the aggregated data from two to twelve panels, all of the plasma cases could be correctly classified.

[0193] Compared with the manual method for performing digital PCR, this microfluidics-based method is much more rapid and less labor-intensive. The whole process could be completed in two and a half hours.

B. Digital RNA-SNP analysis for the prenatal detection of trisomy 18

[0194] In this example, we used a digital PCR-based allelic discrimination assay on *serpin peptidase inhibitor clade B (ovalbumin) member 2* (*SERPINB2*) mRNA, a placenta-expressed transcript on chromosome 18, to detect an imbalance in the ratio of polymorphic alleles for trisomy 18 fetuses. Extraction of DNA and RNA from placental tissue samples was performed using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) and the TRIzol reagent (Invitrogen, Carlsbad, CA), respectively, as described in the manufacturers' protocols. The extracted placental RNA samples were subjected to DNase I treatment (Invitrogen) for removal of contaminating genomic DNA. Genotyping of the rs6098 SNP on

the SERPINB2 gene was carried out in placental tissue DNA samples with a Homogenous MassEXTEND (hME) assay using the MassARRAY Compact (Sequenom, San Diego) as previously described.

[0195] Reverse transcription for the SERPINB2 transcript was performed on the placental tissue RNA samples with a gene-specific primer 5′-CGCAGACTTCTCACCAAACA-3′ (Integrated DNA Technologies, Coralville, IA) using the ThermoScript reverse transcriptase (Invitrogen). All cDNA samples were diluted to a concentration such that the subsequent digital PCR amplification could be performed at an average concentration of one template molecule per reaction well. Digital PCR was set up using the TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA) and the BiomarkTM PCR Reagents (Fluidigm, San Francisco). The forward primer 5′- CTCAGCTCTGCAATCAATGC-3′ (Integrated DNA Technologies) and the reverse primer (identical to the gene-specific primer used for reverse transcription) were used at a concentration of 600 nM. The two TaqMan probes targeting the A or G allele of the rs6098 SNP on the SERPINB2 sequence were 5′-(FAM)CCACAGGGAATTATTT (MGBNFQ)-3′ and 5′-(FAM)CCACAGGGGATTATTT(MGBNFQ)-3′ (Applied Biosystems). FAM is 6-carboxyfluorescein and MGBNFQ is a minor groove-binding nonfluorescent quencher, and were used at concentrations of 300 nM and 500 nM, respectively. Each sample-reagent mix was partitioned into 765 reaction wells on a BiomarkTM 12.765 Digital Array using the NanoflexTM IFC Controller (Fluidigm). After partitioning, the array was placed in the BiomarkTM Real-time PCR System (Fluidigm) for thermal amplification and fluorescence detection. The reaction was initiated at 50oC for 2 min and continued at 95oC for 5 min followed by 45 cycles of 95oC for 15 sec and 59 oC for 1 min. After amplification, the number of informative wells (one that was positive for either the A or G allele only) and the number of wells positive for both alleles were counted and subjected to sequential probability ratio test (SPRT) analysis.

[0196] For a heterozygous euploid fetus, the A and G alleles should be equally represented (1:1) in the fetal genome, whereas for trisomy 18, there would be an additional copy of one allele thus giving a ratio of 2:1 in the fetal genome. A series of SPRT curves were generated for interpretation of different samples. These curves illustrate the expected proportion of informative wells positive for the overrepresented allele P_r (y-axis) for a given total number of informative wells (x-axis) needed for classification. For each sample, the experimentally derived P_r was compared with the expected P_r value. Samples above the upper curve were

classified as trisomy 18, whereas those below the bottom curve were classified as euploid. The area between the two curves is the unclassifiable region.

[0197] The feasibility of digital RNA-SNP analysis for the detection of fetal trisomy 18 was demonstrated by using the rs6098 SNP on the SERPINB2 gene. Placental tissue DNA samples from subjects with euploid and trisomy 18 fetuses were first genotyped by mass spectrometry for identifying heterozygous cases. Nine euploid and three trisomy 18 heterozygous placentas were found and subjected to digital RNA-SNP analysis. For each sample, P_r and m_r were calculated, and the SPRT curve for this m_r value was used for disease classification. As shown in FIG. 16A, all samples were correctly classified. The P_r values for trisomy 18 placentas were above the unclassifiable region, whereas those for euploid placentas fell below this region.

[0198] Samples with SPRT curves based on $m_r = 0.1$, 0.2, and 0.3 are illustrated in FIG. 16B. These data suggest that the digital RNA-SNP method is a valuable diagnostic tool for trisomy 18 pregnancies. The two curves represent the boundaries for the unclassifiable region. Samples with data points above the upper curve were classified as aneuploid, whereas those with data points below the bottom curve were classified as euploid.

C. Digital RCD analysis

[0199] This example demonstrates the performance of digital RCD analysis using microfluidics-based digital PCR. A variant of this approach is illustrated here, by way of example but not by way of limitation to, using a Fluidigm BioMark™ System. This system can perform over 9000 digital PCRs per run.

[0200] Placental tissues, maternal blood cell and amniotic fluid samples were obtained from pregnant women carrying euploid or trisomy 21 (T21) fetuses. Placental DNA from 10 T21 and 10 euploid cases was each mixed with an equal amount of euploid maternal blood cell DNA, thus producing twenty 50% DNA mixtures. To ensure accurate fetal proportion in the mixture samples, the extracted DNA was first quantified by optical density (OD) measurement at 260 nm. They were then digitally quantified by the BioMark™ System (Fluidigm) using the 12.765 Digital Arrays (Fluidigm). The assay for quantifying the samples was the same as described below except that only the probe for the reference chromosome was used.

[0201] The chromosome dosage in the 50% DNA mixtures and amniotic fluid samples was determined by digital PCR analysis of a nonpolymorphic chromosome 21 locus relative to one located on chromosome 1. A 101-bp amplicon of a pair of paralogous loci on chromosome 21 and 1 was first co-amplified by forward primer 5'-GTTGTTCTGCAAAAAACCTTCGA-3' and reverse primer 5'-CTTGGCCAGAAATACTTCATTACCATAT-3'. Two chromosome-specific TaqMan probes were designed to distinguish between the chromosome 21 and 1 paralogs, and their sequences were 5'-(FAM)TACCTCCATAATGAGTAAA(MGBNFQ)-3' and 5'-(VIC)CGTACCTCTGTAATGTGTAA(MGBNFQ)-3', respectively. The use of paralogous loci was used here by way of example only. In other words, non-paralogous loci could also be used for such analysis.

[0202] In order to demonstrate the use of the digital RCD approach to detect trisomy 18 (T18), another assay targeting paralogous sequences on chromosome 21 and 18 was designed. A 128-bp amplicon of the paralogous loci on chromosome 21 and 18 was first co-amplified by forward primer 5'-GTACAGAAACCACAAACTGATCGG-3' and reverse primer 5'-GTCCAGGCTGTGGGCCT-3'. Two chromosome-specific TaqMan probes were designed to distinguish between the chromosome 21 and 18 paralogs, and their sequences were 5'-(FAM)AAGAGGCGAGGCAA(MGBNFQ)-3' and 5'-(VIC)AAGAGGACAGGCAAC(MGBNFQ)-3', respectively. The use of paralogous loci was used here by way of example only. In other words, non-paralogous loci could also be used for such analysis.

[0203] All experiments were carried out on the BioMark™ System (Fluidigm) using the 12.765 Digital Arrays (Fluidigm). The reaction for one panel was set up using 2X TaqMan Universal PCR Master Mix (Applied Biosystems) in a reaction volume of 10 μL. Each reaction contained 1X TaqMan Universal PCR Master Mix, 900 nM of each primer, 125 nM of each probe and 3.5 μL of a 50% placental/maternal blood cell DNA sample. The sample/assay mixture was loaded into the Digital Array by the NanoFlex™ IFC controller (Fluidigm). The reaction was carried out on the BioMark™ System for detection. The reaction was initiated at 50°C for 2 min, followed by 95°C for 10 min and 40 cycles of 95°C for 15s and 57°C for 1 min.

[0204] The euploid and T21 50% placental/maternal blood cell DNA samples were analyzed on the digital arrays with the chr21/chr1 assay. For each sample, the number of

informative wells, comprising the ones positive for only chr21 or chr1 markers (but not both), was counted. The proportion of the overrepresented marker among all the informative wells (P_r) was determined. SPRT curves appropriate for the exact average reference template concentration per well (m_r) for any one of the digital PCR panels were applied to determine if the experimentally-obtained P_r indicated a euploid or T21 sample. Data were aggregated from extra panels for samples which remained unclassified until a decision could be made. As shown in FIG. 17, all 50% placental/maternal blood cell DNA samples were correctly classified using this approach with data ranging from one to four panels needed. A SPRT curve was also plotted to show the decision boundaries for correct classification, as shown in FIG. 18.

[0205] We further applied the RCD analysis on amniotic fluid samples obtained from 23 women carrying euploid fetuses and 6 women carrying T21 fetuses. Each sample was analyzed in a single 765-well reaction panel with the chr21/chr1 assay. FIG. 19 shows the SPRT classification summary. As shown in FIG. 19, all the 29 samples were classified correctly. The digital RCD method is thus an alternative approach for the detection of trisomies using microsatellite (Levett LJ, et al. A large-scale evaluation of amnio-PCR for the rapid prenatal diagnosis of fetal trisomy. Ultrasound Obstet Gynecol 2001; 17: 115-8) or single nucleotide polymorphism (SNP) (Tsui NB, et al. Detection of trisomy 21 by quantitative mass spectrometric analysis of single-nucleotide polymorphisms. Clin Chem 2005; 51: 2358-62) markers or real-time non-digital PCR (Zimmermann B, et al. Novel real-time quantitative PCR test for trisomy 21. Clin Chem 2002; 48: 362-3) in miscellaneous sample types used for prenatal diagnosis, such as amniotic fluid and chorionic villus biopsies.

[0206] In an attempt to detect T18 cases, we applied the chr21/chr18 assay on 3 euploid and 5 T18 placental DNA samples. The proportion of the overrepresented marker among all the informative wells (P_r) was calculated. All of them were classified correctly except one T18 case was misclassified as euploid. The results were summarized in FIG. 20.

V. USING MULTIPLEX DIGITAL RCD ASSAYS ON MASS-SPECTROMETRIC PLATFORM

[0207] The number of plasma DNA molecules that are present per unit volume of maternal plasma is limited (Lo YMD. *et al.* 1998 *Am J Hum Genet* 62, 768-7758). For example, in early pregnancy, the median maternal plasma concentration of an autosomal locus, the *β-globin* gene, has been shown to be 986 copies/mL, with contributions from both the fetus

and mother (Lo YMD. *et al.* 1998 *Am J Hum Genet* 62, 768-7758). To capture 7,680 molecules, DNA extracted from some 8 mL of maternal plasma would be needed. This volume of plasma, obtainable from ~15 mL of maternal blood, is at the limit of routine practice. However, we envision that multiple sets of chr21 and reference chromosome targets can be combined for digital RCD analysis. For 5 pairs of chr21 and reference chromosome targets, just 1.6 mL of maternal plasma would be needed to provide the number of template molecules needed for analysis. Multiplex single molecule PCR could be performed. The robustness of such multiplex single molecule analysis has been demonstrated previously for single molecule haplotyping (Ding, C. and Cantor, CR. 2003 *Proc Natl Acad Sci U S A* 100, 7449-7453).

[0208] In one example, placental tissues and maternal blood cell samples were obtained from pregnant women carrying euploid or trisomy 21 (T21) fetuses. 5 euploid and 5 T21 placental DNA samples were each mixed with equal proportions of maternal blood cell DNA to produce 10 DNA mixtures mimicking plasma samples with 50% fetal DNA. To ensure accurate fetal proportion in the mixture samples, the extracted DNA was first quantified by optical density (OD) measurement at 260 nm. They were then digitally quantified by real-time PCR in 384-well format. The assay for quantifying the samples was the same as described in the previous example of digital RCD analysis.

[0209] The chromosome dosage in the 50% mix was determined by digital PCR analysis of a nonpolymorphic chromosome 21 locus relative to one located on chromosome 1. The method is called Digital Relative Chromosome Dosage (RCD) analysis. A 121-bp amplicon (inclusive of a 10-mer on each primer) of a pair of paralogous loci on chromosome 21 and 1 was co-amplified by forward primer 5'-ACGTTGGATGGTTGTTCTGCAAAAAACCTTCGA-3' and reverse primer 5'-ACGTTGGATGCTTGGCCAGAAATACTTCATTACCATAT-3'. An extension primer which targets the base differences between chromosome 21 and chromosome 1 was designed, and its sequence is 5'-CTCATCCTCACTTCGTACCTC-3'.

[0210] In order to demonstrate the utility of multiplexing digital PCR assays to detect T21 cases, another digital RCD assay targeting paralogous sequences on chromosome 21 and 18 was designed. A 148-bp amplicon (inclusive of a 10-mer on each primer) of the paralogous loci on chromosome 21 and 18 was co-amplified by forward primer 5'-ACGTTGGATGGTACAGAAACCACAAACTGATCGG-3' and reverse primer 5'-

ACGTTGGATGGTCCAGGCTGTGGGCCT-3'. An extension primer which targets the base differences between chromosome 21 and chromosome 18 was designed, and its sequence is 5'-ACAAAAGGGGGAAGAGG-3'.

[0211] Multiplex digital RCD analysis was performed using primer extension protocol. PCR reaction was set up using GeneAmp PCR Core Reagent Kit (Applied Biosystems) in a reaction volume of 5 μL. Each reaction contained 1X Buffer II, 2 mM $MgCl_2$, 200 μM dNTP mix, 0.2 U AmpliTaq Gold, 200 nM of each of the 4 primers and the 50% DNA mix. The assay/sample mixture was dispensed into 384-well PCR plate and the reaction was initiated at 50°C for 2 min, followed by 95°C for 10 min and 40 cycles of 95°C for 15s and 57°C for 1 min.

[0212] PCR products were subjected to shrimp alkaline phosphatase (SAP) treatment to remove unincorporated dNTPs. The mixture was incubated at 37°C for 40 min followed by 85°C for 5 min. Primer extension reaction was then carried out. In brief, 771 nM of extension primer from chr21/chr1 assay, 1.54 μM of extension primer from chr21/chr18 assay, 0.67 U Thermosequenase (Sequenom), and 64 μM each of ddCTP, ddGTP, dATP and dTTP in an extension cocktail were added to the SAP-treated PCR products. The reaction conditions were 94°C for 2 min, followed by 94°C for 5 s, 50°C for 5 s, and 72°C for 5 s for 80 cycles. 16 μL of water and 3 mg of the Clean Resin (Sequenom) were added to the extension products for a final clean up. The mixtures were mixed in a rotator for 20 to 30 min, followed by a centrifugation step at 361 g for 5 min. Fifteen to 25 nL of the final products were dispensed onto a SpectroCHIP (Sequenom) by a MassARRAY Nanodispenser S (Sequenom). Data acquisition from the SpectroCHIP was done in the MassARRAY Analyzer Compact Mass Spectrometer (Sequenom). Mass data were imported into the MassARRAY Typer (Sequenom) software for analysis.

[0213] The five euploid and five T21 50% placental/maternal DNA samples were analyzed with the duplex RCD assay. For each sample, the number of informative wells from individual assay, comprising the ones positive for only chr21 or chr1 or chr18 markers, was counted. The proportion of the chr21 marker among all the informative wells (P_r) was calculated separately for each RCD assay. Sequential probability ratio test (SPRT) was then applied to determine if the P_r indicated a euploid or T21 sample. By doing so, the number of wells required was reduced as each plate was counted twice.

[0214] The chr21/chr1 assay was usually applied first. If the sample remained unclassified, then the values from the chr21/chr18 assay would be added for further calculations. Extra plates were used for samples which remained unclassified until a decision could be made. As shown in FIG. 21, all euploid 50% mix samples were correctly classified using a single 384-well plate. Several T21 cases required 2 or more plates for correct classification. If only one assay was used, a greater number of plates would be needed to attain the number of informative wells required when classification was achieved. For example, data for case N0230 was unclassifiable when either of the RCD assays was used alone. However, correct classification was achieved when data from the two assays were combined. If the duplex RCD assays were not used, additional plates of analyses would be needed. We would expect a further reduction of well number with a higher level of multiplexing of assays.

[0215] In another example, we developed a 4-plex assay targeting 4 different amplicons on chromosome 21 and their corresponding paralogous partners located on other non-chromosome-21 autosomes. This 4-plex assay was used in digital RCD analysis followed by SPRT classification of samples from euploid and trisomy 21 pregnancies. DNA extractions from placental samples were performed using the QIAamp tissue kit (Qiagen, Hilden, Germany).

[0216] All placental and maternal buffy coat DNA samples used in this study were first quantified by the NanoDrop spectrophotometer (NanoDrop Technology, Wilmington, DE). The DNA concentration was converted to genome equivalent (GE)/μL using a conversion of 6.6 pg/cell. The amount of DNA corresponding to approximately one template per well was determined by serially diluting the DNA samples. Under such a condition, we would expect approximately 37% of the wells to show negative amplification. In multiplex digital RCD analysis, 4 sets of paralogous sequence targets were selected: the paralogous loci on chromosome 21 and 1 were co-amplified by the forward primer 5'-ACGTTGGATGTTGATGAAGTCTCATCTCTACTTCG -3' and the reverse primer 5'-ACGTTGGATGCAATAAGCTTGGCCAGAAATACT -3', resulting in an amplicon of 81 bp. The paralogous loci on chromosome 21 and 7 were co-amplified by the forward primer 5'- ACGTTGGATGGAATTTAAGCTAAATCAGCCTGAACTG-3' and the reverse primer 5'- ACGTTGGATGGTTTCTCATAGTTCATCGTAGGCTTAT -3', resulting in an amplicon of 82 bp. The paralogous loci on chromosome 21 and 2 were co-amplified by the forward primer 5'- ACGTTGGATGTCAGGCAGGGTTCTATGCAG -3' and the reverse primer 5'-ACGTTGGATGAGGCGGCTTCCTGGCTCTTG -3', resulting in an amplicon of 101 bp.

The paralogous loci on chromosome 21 and 6 were co-amplified by the forward primer 5'-ACGTTGGATGGCTCGTCTCAGGCTCGTAGTT -3' and the reverse primer 5'-ACGTTGGATGTTTCTTCGAGCCCTTCTTGG -3', resulting in an amplicon of 102 bp. Each reaction contained 10X buffer II (Applied Biosystems), $MgCl_2$ and 100 nM of each primer. The total reaction volume was 5 μL/well. The reaction was initiated at 95°C for 5 min, followed by 45 cycles of 95°C for 30 sec, 62 °C for 30 sec and 72°C for 30 sec, and a final extension at 72 °C for 7 min. All conventional PCR amplifications were carried out on a GeneAmp PCR System 9700 (Applied Biosystems). The unincorporated nucleotides were deactivated by shrimp alkaline phosphatase (SAP) treatment. Each reaction contained 10X SAP buffer (Sequenom) and SAP enzyme (Sequenom). 2 μL of SAP mix was added to each PCR. The SAP reaction was incubated at 37 °C for 40 min and 85 °C for 5 min. After the SAP treatment, primer extension reaction was carried out on the PCR products using the iPLEX Gold kit (Sequenom). . The paralogous sequence mismatches (PSMs) on the paralogous loci on chromosomes 21 and 1 were interrogated by the extension primer 5'-GTCTCATCTCTACTTCGTACCTC -3'. The PSMs on the paralogous loci on chromosomes 21 and 7 were interrogated by the extension primer 5'-TTTTACGCTGTCCCCATTT-3'. The PSMs on the paralogous loci on chromosomes 21 and 2 were interrogated by the extension primer 5'-GGTCTATGCAGGAGCCGAC-3'. The PSMs on the paralogous loci on chromosomes 21 and 6 were interrogated by the extension primer 5'-TGGGCGCGGGAGCGGACTTCGCTGG-3'. Each reaction contained 10X iPLEX buffer (Sequenom), iPLEX termination mix (Sequenom), iPLEX enzyme (Sequenom) and 343 nM of each extension primer, except for the extension primer for PSMs on chromosomes 21 and 6 which was used at 1.03 μM. 2 μL of iPLEX mix was added to 5 μL of PCR product. The iPLEX reaction was cycled according to a 200-short-cycle program. Briefly, the samples were first denatured at 94 °C for 35 sec, followed by annealing at 52° C for 5 sec and extension at 80°C for 5 sec. The annealing and extension cycle was repeated four more times for a total of five cycles and then looped back to a 94° C denaturing step for 5 sec, after which was the 5-cycle annealing and extension loop again. The five annealing and extension cycles with the single denaturing step were repeated 39 times for a total of 40. A final extension at 72° C for 3 min was performed. The iPLEX reaction products were diluted with 16 μL water and desalted by 6 mg resin for each PCR. The 384-well plate was centrifuged at 1600 *g* for 3 min before dispensing onto the SpectroCHIP (Sequenom) and the matrix-assisted laser

desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry MS Analysis (Sequenom).

[0217] The number of wells which were positive for only chromosome 21 or only the reference chromosome for each of the four assays was independently recorded. For each assay, the Poisson corrected numbers of molecules for chromosome 21 and the reference chromosome were calculated. The sum of the Poisson corrected number of molecules for chromosome 21 as well as the sum of the Poisson corrected number of reference chromosomes from all four assays were calculated and deemed as the informative counts for the 4-plex assay. The P_r value was the chromosome 21 count for the 4-plex assay divided by the sum of the chromosome 21 and reference chromosome counts for the 4-plex assay. The experimentally derived P_r values were subjected to SPRT analysis. One or more 384-well plates would be analyzed until disease classification was possible by SPRT. A total of two 50% euploid placental genomic DNA/50% maternal buffy coat DNA mix and two 50% trisomy 21 placental genomic DNA/50% maternal buffy coat DNA mix were analyzed.

[0218] The experimentally derived P_r value would be compared with the expected value of P_r to test the null or alternative hypotheses. Alternatively, neither the null or alternative hypothesis could be accepted if the P_r for the given number of informative counts has not yet reached the required level of statistical confidence for disease classification. These cases were deemed unclassifiable until more data were available.

[0219] The results and the SPRT classification of each sample are tabulated in FIGS. 22A and 22B. The two euploid samples required 2 and 5 384-well multiplex digital RCD analyses before SPRT classification could be reached. Data from none of the individual member of the 4-plex assay allowed disease classificaiton by SPRT. Both trisomy 21 samples were each correctly classified with just one 384-well multiplex digital RCD analysis. Similarly, data from none of the individual member of the 4-plex assay allowed disease classificaiton by SPRT. However, the composite counts from the 4-plex assay allowed correct SPRT classification. These data illustrated that by using multiplex digital RCD, the effective number of informative counts were substantially increased for a given number of digital PCR analyses performed as compared to the use of a single-plex digital RCD assay.

VI. USING DIGITAL EPIGENETIC RELATIVE CHROMOSOME DOSAGE

[0220] Here we outline an approach called digital epigenetic relative chromosome dosage (digital ERCD) in which epigenetic markers exhibiting a fetal-specific DNA methylation pattern, or other epigenetic changes, on a chromosome involved in a chromosomal aneuploidy (e.g. chromosome 21 in trisomy 21) and on a reference chromosome, are subjected to digital PCR analysis. The ratio of the number of wells positive for the chromosome 21 epigenetic marker to that positive for the reference chromosome epigenetic marker in plasma DNA extracted from pregnant women bearing normal fetuses will give us the reference range. The ratio will be expected to be increased if the fetus has trisomy 21. It is obvious to those of skill in the art that more than one chromosome 21 markers and more than one reference chromosome markers could be used in this analysis.

[0221] One example of a gene on chromosome 21 which exhibits a fetus (placenta)-specific methylation pattern is the *Holocarboxylase synthetase (HLCS)* gene. *HLCS* is hypermethylated in the placenta, but hypomethylated in maternal blood cells; and is covered in U.S. patent application 11/784499, which is incorporated herein by reference. One example of a gene on a reference chromosome which exhibits a fetus (placenta)-specific methylation pattern is the *RASSF1A* gene on chromosome 3 [10]. *RASSF1A* is hypermethylated in the placenta but is hypomethylated in maternal blood cells, see U.S. patent application 11/784501, which is incorporated herein by reference.

[0222] In the application of hypermethylated *HLCS* and hypermethylated *RASSF1A* to digital PCR detection of trisomy 21 in a fetus using maternal plasma, maternal peripheral blood is first collected. Then the blood is subjected to centrifugation and the plasma is harvested. DNA from the plasma is then extracted using techniques well-known to those of skill in the art, such as using a QIAamp Blood kit (Qiagen). The plasma DNA is then subjected to digestion using one or more methylation-sensitive restriction enzymes, such as *Hpa*II and *Bst*UI. These methylation-sensitive restriction enzyme(s) will cut the maternal, nonmethylated versions of these genes, while leaving the fetal hypermethylated sequences intact. The digested plasma DNA sample is then diluted to an extent that on average approximately 0.2 to 1 molecule of either the restriction enzyme treated but intact *HLCS* or *RASSF1A* sequences will be detected per reaction well. Two real-time PCR systems will be used to amplify the diluted DNA, one with two primers and one TaqMan probe specific to the *HLCS* gene, encompassing the region that will be cut by the restriction enzyme(s) if the

sequence in unmethylated; and the other one towards the *RASSF1A* gene, similarly with two primers and one TaqMan probe. With regard to the latter *RASSF1A* primer/probe set, one example has been described by Chan et al 2006, Clin Chem 52, 2211-2218. The TaqMan probes towards the *HLCS* and *RASSF1A* targets will have different fluorescent reporters, such as FAM and VIC, respectively. A 384-well plate is then used to perform the digital PCR experiment. The number of wells scored positive for just *HLCS* and those scored positive for just *RASSF1A* will be counted, and a ratio of these counts will be taken. The *HLCS*:*RASSF1A* ratio will be expected to be higher for maternal plasma taken from a pregnant woman carrying a trisomy 21 fetus, when compared with one carrying a normal euploid fetus. The degree of overrepresentation will be dependent on the average reference template concentration per well in the digital PCR run.

[0223] Other methods for scoring these results will be possible, for the example the counting of the number of wells positive for *HLCS*, irrespective of the concurrent positivity for *RASSF1A*; and vice versa for *RASSF1A*, irrespective of the concurrent positivity for *HLCS*. Furthermore, in replacement of calculating the ratio, either the total number, or the difference in the *HLCS* and *RASSF1A* counts could be used to indicate the trisomy 21 status of a fetus.

[0224] Apart from doing the digital PCR in plates, it will also be obvious to those of skill in the art that other variants of digital PCR can be used, e.g. microfluidics chips, nanoliter PCR microplate systems, emulsion PCR, polony PCR and rolling-circle amplification, primer extension and mass spectrometry, etc. These variants of digital PCR are named by way of examples, and not as limitations.

[0225] Apart from real-time PCR, it will also be obvious to those of skill in the art that methods such as mass spectrometry can be used to score the digital PCR results.

[0226] Apart from using methylation-sensitive restriction enzymes to differentiate the fetal and maternal versions of *HLCS* and *RASSF1A*, it will be obvious to those of skill in the art that other methods for ascertaining the methylation status would also be applicable, e.g. bisulfite modification, methylation-specific PCR, immunoprecipitation using antibody to methylated cytosine, mass spectrometry, etc.

[0227] It will also be obvious to those of skill in the art that the approach illustrated in this example and other examples in this patent application can be used in the other bodily fluids in

which fetal DNA may be found, including maternal urine, amniotic fluid, transcervical washings, chorionic villus, maternal saliva, etc.

VII. MASSIVELY PARALLEL GENOMIC SEQUENCING USING EMULSION PCR AND OTHER STRATEGIES

[0228] Here we shall describe another example whereby a digital readout of nucleic acid molecules can be used for the detection of fetal chromosomal aneuploidies, e.g. trisomy 21, in maternal plasma. Fetal chromosomal aneuploidy results from abnormal dose(s) of a chromosome or chromosomal region. It is desirable that noninvasive tests have high sensitivity and specificity to minimize false diagnoses. However, fetal DNA is present in low absolute concentration and represent a minor portion of all DNA sequences in maternal plasma and serum. Hence, the number of digital PCR sampling targeting specific gene loci cannot be increased infinitely within the same biological specimen. Hence, the analysis of multiple sets of specific target loci may be used to increase the amount of data that could be obtained from a specimen without increasing the number of digital PCR sampling performed.

[0229] Accordingly, embodiments allow the noninvasive detection of fetal chromosomal aneuploidy by maximizing the amount of genetic information that could be inferred from the limited amount of fetal nucleic acids which exist as a minor population in a biological sample containing maternal background nucleic acids. In one aspect, the amount of genetic information obtained is sufficient for accurate diagnosis yet not overly excessive so as to contain costs and the amount of input biological sample required.

[0230] Massively parallel sequencing, such as that achievable on the 454 platform (Roche) (Margulies, M. *et al.* 2005 *Nature* 437, 376-380), Illumina Genome Analyzer (or Solexa platform) or SOLiD System (Applied Biosystems) or the Helicos True Single Molecule DNA sequencing technology (Harris TD et al. 2008 Science, 320, 106-109), the single molecule, real-time (SMRT^{TM}) technology of Pacific Biosciences, and nanopore sequencing (Soni GV and Meller A. 2007 Clin Chem 53: 1996-2001), allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion (Dear Brief Funct Genomic Proteomic 2003; 1: 397-416). Each of these platforms sequences clonally expanded or even non-amplified single molecules of nucleic acid fragments.

[0231] As a high number of sequencing reads, in the order of hundred thousands to millions or even possibly hundreds of millions or billions, are generated from each sample in each run,

the resultant sequenced reads form a representative profile of the mix of nucleic acid species in the original specimen. For example, the haplotype, trascriptome and methylation profiles of the sequenced reads resemble those of the original specimen (Brenner et al Nat Biotech 2000; 18: 630-634; Taylor et al Cancer Res 2007; 67: 8511-8518). Due to the large sampling of sequences from each specimen, the number of identical sequences, such as that generated from the sequencing of a nucleic acid pool at several folds of coverage or high redundancy, is also a good quantitative representation of the count of a particular nucleic acid species or locus in the original sample.

[0232] In one embodiment, random sequencing is performed on DNA fragments that are present in the plasma of a pregnant woman, and one obtains genomic sequences which would originally have come from either the fetus or the mother. Random sequencing involves sampling (sequencing) a random portion of the nucleic acid molecules present in the biological sample. As the sequencing is random, a different subset (fraction) of the nucleic acid molecules (and thus the genome) may be sequenced in each analysis. Embodiments will work even when this subset varies from sample to sample and from analysis to analysis, which may occur even using the same sample. Examples of the fraction are about 0.1%, 0.5%, or 1% of the genome. In other embodiments, the fraction is at least any one of these values.

[0233] A bioinformatics procedure may then be used to locate each of these DNA sequences to the human genome. It is possible that a proportion of such sequences will be discarded from subsequent analysis because they are present in the repeat regions of the human genome, or in regions subjected to inter-individual variations, e.g. copy number variations. An amount of the chromosome of interest and of one or more other chromosomes may thus be determined.

[0234] In one embodiment, a parameter (e.g. a fractional representation) of a chromosome potentially involved in a chromosomal aneuploidy, e.g. chromosome 21 or chromosome 18 or chromosome 13, may then be calculated from the results of the bioinformatics procedure. The fractional representation may be obtained based on an amount of all of the sequences (e.g. some measure of all of the chromosomes) or a particular subset of chromosomes (e.g. just one other chromosome than the one being tested.)

[0235] This fractional representation is then compared to a reference range established in pregnancies involving normal (i.e. euploid) fetuses. It is possible that in some variants of the

procedure, the reference range would be adjusted in accordance with the fractional concentration of fetal DNA (f) in a particular maternal plasma sample. The value of f can be determined from the sequencing dataset, e.g. using sequences mappable to the Y chromosome if the fetus is male. The value of f may also be determined in a separate analysis, e.g. using fetal epigenetic markers (Chan KCA et al 2006 Clin Chem 52, 2211-8) or from the analysis of single nucleotide polymorphisms.

[0236] In one aspect, even when a pool of nucleic acids in a specimen is sequenced at <100% genomic coverage, and among the proportion of captured nucleic acid molecules, most of each nucleic acid species is only sequenced once, dosage imbalance of a particular gene locus or chromosome can also be quantitatively determined. In other words, the dosage imbalance of the gene locus or chromosome is inferred from the percentage representation of the said locus among all mappable sequenced tags of the specimen.

[0237] In one aspect for the massively parallel genomic sequencing approach, representative data from all of the chromosomes may be generated at the same time. The origin of a particular fragment is not selected ahead of time. The sequencing is done at random and then a database search may be performed to see where a particular fragment is coming from. This is contrasted from situations when a specific fragment from chromosome 21 and another one from chromosome 1 are amplified.

[0238] In one example, a proportion of such sequences would be from the chromosome involved in an aneuploidy such as chromosome 21 in this illustrative example. Yet other sequences from such a sequencing exercise would be derived from the other chromosomes. By taking into account of the relative size of chromosome 21 compared with the other chromosomes, one could obtain a normalized frequency, within a reference range, of chromosome 21-specific sequences from such a sequencing exercise. If the fetus has trisomy 21, then the normalized frequency of chromosome 21-derived sequences from such a sequencing exercise will increase, thus allowing the detection of trisomy 21. The degree of change in the normalized frequency will be dependent on the fractional concentration of fetal nucleic acids in the analyzed sample.

[0239] In one embodiment, we used the Illumina Genome Analyzer for single-end sequencing of human genomic DNA and human plasma DNA samples. The Illumina Genome Analyzer sequences clonally-expanded single DNA molecules captured on a solid surface termed a flow cell. Each flow cell has 8 lanes for the sequencing of 8 individual

specimens or pools of specimens. Each lane is capable of generating ~ 200Mb of sequence which is only a fraction of the 3 billion basepairs of sequences in the human genome. Each genomic DNA or plasma DNA sample was sequenced using one lane of a flow cell. The short sequence tags generated were aligned to the human reference genome and the chromosomal origin was noted. The total number of individual sequenced tags aligned to each chromosome were tabulated and compared with the relative size of each chromosome as expected from the reference human genome or non-disease representative specimens. Chromosome gains or losses were then identified.

[0240] The described approach is only one exemplification of the presently described gene/chromosome dosage strategy. Alternatively, paired end sequencing could be performed. Instead of comparing the length of the sequenced fragments from that expected in the reference genome as described by Campbell et al (Nat Genet 2008; 40: 722-729), the number of aligned sequenced tags were counted and sorted according to chromosomal location. Gains or losses of chromosomal regions or whole chromosomes were determined by comparing the tag counts with the expected chromosome size in the reference genome or that of a non-disease representative specimen

[0241] In another embodiment, the fraction of the nucleic acid pool that is sequenced in a run is further sub-selected prior to sequencing. For example, hybridization based techniques such as oligonucleotide array could be used to first sub-select for nucleic acid sequences from certain chromosomes, e.g. a potentially aneuploid chromosome and other chromosome(s) not involved in the aneuploidy tested. Another example is that a certain sub-population of nucleic acid sequences from the sample pool is sub-selected or enriched prior to sequencing. For example, it has been reported that fetal DNA molecules in maternal plasma are comprised of shorter fragments than the maternal background DNA molecules (Chan et al Clin Chem 2004; 50: 88-92). Thus, one may use one or more methods known to those of skill in the art to fractionate the nucleic acid sequences in the sample according to molecule size, e.g. by gel electrophoresis or size exclusion columns or by microfluidics-based approach. Yet, alternatively, in the example of analyzing cell-free fetal DNA in maternal plasma, the fetal nucleic acid portion could be enriched by a method that suppresses the maternal background, such as by the addition of formaldehyde (Dhallan et al JAMA 2004; 291: 1114-9).

[0242] Other single molecule sequencing strategies such as that by the Roche 454 platform, the Applied Biosystems SOLiD platform, the the Helicos True Single Molecule DNA

sequencing technology, the single molecule, real-time (SMRT™) technology of Pacific Biosciences, and nanopore sequencing could similarly be used in this application.

[0243] Examples of results and a further discussion (e.g. for sequencing and calculating parameters) may be found in concurrently filed application "DIAGNOSING FETAL CHROMOSOMAL ANEUPLOIDY USING GENOMIC SEQUENCING," (Attorney Docket No. 016285-005220US), which is incorporated by reference. Note that methods described herein for determining a cutoff value may be implemented when the reaction is sequencing, e.g. as described in this section.

[0244] The determination of the fractional concentration of fetal DNA in maternal plasma can also be done separate to the sequencing run. For example, the Y chromosome DNA concentration could be pre-determined using real-time PCR, microfluidics PCR or mass spectrometry. In fact, fetal DNA concentration could be determined using loci other than the Y chromosome and applicable to female fetuses. For example, Chan et al showed that fetal-derived methylated *RASSF1A* sequences would be detected in the plasma of pregnant women in the background of maternally derived unmethylated *RASSF1A* sequences (Chan et al, Clin Chem 2006;52:2211-8). The fractional fetal DNA concentration can thus be determined by dividing the amount of methylated *RASSF1A* sequences by the amount of total *RASSF1A* (methylated and unmethylated) sequences.

[0245] It is expected that maternal plasma would be preferred over maternal serum for practicing our invention because DNA is released from the maternal blood cells during blood clotting. Thus, if serum is used, it is expected that the fractional concentration of fetal DNA will be lower in maternal plasma than maternal serum. In other words, if maternal serum is used, it is expected that more sequences would need to be generated for fetal chromosomal aneuploidy to be diagnosed, when compared with a plasma sample obtained from the same pregnant woman at the same time.

[0246] Yet another alternative way of determining the fractional concentration of fetal DNA would be through the quantification of polymorphic differences between the pregnant women and the fetus (Dhallan R, et al. 2007 *Lancet*, 369, 474-481). An example of this method would be to target polymorphic sites at which the pregnant woman is homozygous and the fetus is heterozygous. The amount of fetal-specific allele can be compared with the amount of the common allele to determine the fractional concentration of fetal DNA.

[0247] In contrast to the existing techniques for detecting chromosomal aberrations, including comparative genomic hybridization, microarray comparative genomic hybridization, quantitative real-time polymerase chain reaction, which detect and quantify one or more specific sequence(s), massively parallel sequencing is not dependent on the detection or analysis of predetermined or a predefined set of DNA sequences. A random representative fraction of DNA molecules from the specimen pool is sequenced. The number of different sequence tags aligned to various chromosomal regions is compared between specimens containing or not containing tumoral DNA. Chromosomal aberrations would be revealed by differences in the number (or percentage) of sequences aligned to any given chromosomal region in the specimens.

[0248] In another example the sequencing technique on plasma cell-free DNA may be used to detect the chromosomal aberrations in the plasma DNA for the detection of a specific cancer. Different cancers have a set of typical chromosomal aberrations. Changes (amplifications and deletions) in multiple chromosomal regions may be used. Thus, there would be an increased proportion of sequences aligned to the amplified regions and a decreased proportion of sequences aligned to decreased regions. The percentage representation per chromosome could be compared with the size for each corresponding chromosome in a reference genome expressed as percentage of genomic representation of any given chromosome in relation to the whole genome. Direct comparisons or comparisons to a reference chromosome may also be used.

VIII. MUTATION DETECTION

[0249] Fetal DNA in maternal plasma exists as a minor population, with an average of 3% to 6% of maternal plasma DNA being contributed by the fetus. Because of this reason, most of the previous work in the field has focused on the detection of DNA targets which the fetus has inherited from the father, and which are distinguishable from the majority maternal DNA background in maternal plasma. Examples of such previously detected targets include the *SRY* gene on the Y chromosome (Lo YMD et al. 1998 *Am J Hum Genet,* 62, 768-775) and the *RHD* gene when the mother is RhD-negative (Lo YMD et al. 1998 *N Engl J Med,* 339, 1734-1738.

[0250] For fetal mutation detection, previous strategies using maternal plasma are limited to autosomal dominant conditions in which the father is a carrier, the exclusion of autosomal recessive diseases by direct mutation detection when the father and mother carries different

mutations, or by linkage analysis (Ding C. et al 2004 *Proc Natl Acad Sci U S A* 101, 10762-10767). These previous strategies have significant limitations. For example, for a couple where both the male and female partners carry the same mutation, then it would be impossible to carry out meaningful prenatal diagnosis by direct mutation detection in maternal plasma.

[0251] Such a scenario is illustrated in FIG. 23. In this scenario, there will be three possible fetal genotypes, NN, NM and MM, where N represents the normal allele and M represents the mutant allele. Examples of mutant alleles include those causing cystic fibrosis, beta-thalassemia, alpha-thalassemia, sickle cell anemia, spinal muscular atrophy, congenital adrenal hyperplasia, etc. Other examples of such disorders can be found in the Online Mendelian Inheritance in Man (OMIM) www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM&itool=toolbar. In maternal plasma, most of the DNA will be coming from the mother and would be NM. For any of the three fetal genotypes, there will not be any unique fetal allele which would allow its unique detection in maternal plasma. Thus, the conventional strategy cannot be applied here.

[0252] Embodiments described herein allow handing such scenarios. In the scenario where the mother and fetus are both NM, then the N allele and M allele will be in allelic balance. However, if the mother is NM and the fetus is NN, then there will be allelic imbalance in maternal plasma, with the N allele being overrepresented. On the other hand, if the mother is NM and the fetus is MM, then there will be allelic imbalance in maternal plasma, with the M allele being overrepresented. Thus, for fetal mutation detection, the null hypothesis refers to the absence of allelic imbalance when the fetus is of the NM genotype. The alternative hypothesis refers to the presence of allelic imbalance and the fetal genotype could be NN or MM depending on whether the N or M allele is overrepresented.

[0253] The presence or absence of allelic imbalance can be determined by digital PCR using embodiments described herein. In a first scenario, a particular volume of maternal plasma contains the DNA released from 100 cells, in which 50 are from the mother and 50 are from the fetus. Thus, the fractional concentration of fetal DNA in this volume of plasma is 50%. When the mother is of the genotype NM, then there will be 50 N alleles and 50 M alleles contributed by the mother. If the fetus is of the genotype NM, then there will be 50 N alleles and 50 M alleles contributed by the fetus. Therefore, there will be no allelic imbalance between the N allele and the M allele, a total of 100 copies each. On the other hand, if the

fetus is of the NN genotype, then there will be 100 fetal-derived N alleles in this volume of plasma. Thus, there will be a total of 150 N alleles to 50 M alleles. In other words, there will be allelic imbalance between N and M, with N being overrepresented at a ratio of 3:1 in relation to M.

[0254] In the converse situation, if the fetus is of the MM genotype, then there will be 100 fetal-derived M alleles in this volume of plasma. Thus, there will be 150 M alleles to 50 N alleles. In other words, there will be allelic imbalance between N and M, with M being overrepresented at a ratio of 3:1 in relation to N. Such allelic imbalance can be measured by digital PCR. The allele with the smaller number of positive wells is considered as the reference template. Similar to digital RNA-SNP and digital RCD analyses, the actual distribution of the alleles in the digital PCR experiment would be governed by the Poisson probability density function. Therefore, while the theoretical degree of allelic imbalance in the present scenario is 3:1, the expected degree of allelic imbalance would be dependent on the average template concentration per well during the digital PCR analysis. Thus interpretation cutoffs, such as for SPRT analysis, appropriate for the average reference template concentration per well (m_r) would need to be used for case classification.

[0255] Furthermore, the degree of allelic imbalance that needs to be measured is dependent on the fractional fetal DNA concentration. In contrast to the above example, let's consider a particular volume of maternal plasma contains the DNA released from 100 cells, in which 90 are from the mother and 10 are from the fetus. Thus, the fractional concentration of fetal DNA in this volume of plasma is 10%. When the mother is of the genotype NM, then there will be 90 N alleles and 90 M alleles contributed by the mother. If the fetus is of the genotype NM, then there will be 10 N alleles and 10 M alleles contributed by the fetus. Therefore, there will be no allelic imbalance between the N allele and the M allele, a total of 100 copies each. On the other hand, if the fetus is of the NN genotype, then there will be 20 fetal-derived N alleles in this volume of plasma. Thus, there will be a total of 110 N alleles to 90 M alleles.

[0256] In other words, there will be allelic imbalance between N and M, with N being overrepresented. In the converse situation, if the fetus is of the MM genotype, then there will be 20 fetal-derived M alleles in this volume of plasma. Thus, there will be 110 M alleles to 90 N alleles. In other words, there will be allelic imbalance between N and M, with M being overrepresented. The theoretical degree of allelic imbalance when the fetal DNA fractional concentration is 10% would be 110:90, which is different to the 3:1 ratio when there is 50%

fetal DNA as shown in the above example. Thus interpretation cutoffs, such as for SPRT analysis, appropriate for the fetal DNA fractional concentration would need to be used for case classification.

[0257] Thus, plasma DNA will be extracted. The amount of maternal and fetal DNA in the plasma sample will be quantified, for example by the real-time PCR assays previously established (Lo, *et al.* 1998 *Am J Hum Genet* 62, 768-775) or other types of quantifier well-known to those of skill in the art, e.g. SNP markers (Dhallan R et al. 2007 *Lancet,* 369, 474-481) and fetal epigenetic markers (Chan KCA et al. 2006 *Clin Chem,* 52, 2211-2218). The fetal DNA percentage will be calculated. Then the quantified plasma DNA sample is prepared (e.g. diluted or concentrated) such that during digital PCR analysis, each reaction well will contain an average of one template molecule (can be either the N or M allele). The digital PCR analysis will be carried out using a pair of primers, plus two TaqMan probes, one specific to the N allele, while the other one specific to the M allele. The number of wells which are positive only for M and the number of wells which are positive only for N will be counted. The ratio of these wells will be used to determine if there is evidence of allelic imbalance. Statistical evidence of allelic imbalance can be sought by methods well-known to those of skill in the art, such as using SPRT. In one variant of this analysis, it is also possible to count the number of wells which are positive either for M only or for M and N; as well as to count the number of wells which are positive either for N only or for M and N; and to derive a ratio of these counts. Once again, statistical evidence of allelic imbalance can be sought by methods well-known to those of skill in the art, such as using SPRT.

[0258] The dosage determination of fetal gene mutation, called digital relative mutant dosage (RMD), was validated using female/male (XX/XY) DNA mixtures. Blood cell DNA from a male and a female was each mixed with male DNA, producing samples with XX or XY genotypes in a background of XY at fractional concentrations of 25% and 50%, respectively, as shown in FIG. 24A.

[0259] In addition, blood cell samples were also obtained from 12 male and 12 female subjects. The female blood cell DNA (genotype XX) was each mixed with a 3-fold excess of male blood cell DNA (genotype XY), thus producing 12 DNA mixtures with 25% of DNA with XX genotype in a background of 75% DNA with XY genotype, with results shown in FIG. 24B.

[0260] An aim of the SPRT analysis was to determine the minor genotype present in the background DNA. In DNA mixtures with 25% of XX DNA in a background of 75% XY DNA, the minor allele would be the Y derived from the 75% of DNA. Since 25% of the DNA in the sample was of XX genotype, if there were a total of 200 molecules of DNA in the sample, then 150 molecules would have originated from the XY individual. Hence, the number of Y alleles would be expected to be 75. The number of X alleles contributed by the male proportion of DNA (genotype XY) is also 75. The number of X alleles contributed by the female (genotype XX) is 50 (2 times 25). Therefore, X to Y ratio is 125/75 = (1 +25%) / (1 – 25%) = 5/3.

[0261] For the second part of this study, blood cell samples were obtained from male and female subjects carrying HbE (G → A) and CD41/42 (CTTT/-) mutations on the *beta-globin* gene, i.e., the *hemoglobin, beta* (*HBB*) gene. To mimic maternal plasma samples obtained from heterozygous mothers (MN, where M= mutant and N= wildtype) bearing male fetuses with all possible genotypes (MM, MN or NN), blood cell DNA from males who were either homozygous for the wildtype alleles (NN) or heterozygous (MN) for one of the two mutations was each mixed with a blood cell DNA sample collected from females heterozygous for the same mutation (MN). DNA mixtures at various fractional male/mutant DNA concentrations were thus produced. Blood cell DNA sample from a female homozygous for the CD41/42 deletion (MM) was also used for preparing the DNA mixtures. To ensure an accurate male proportion for the SPRT classification, the fractional male DNA concentration of each DNA mixture was determined using the *ZFY/X* assay.

[0262] The digital *ZFY/X* assay was used for validating the SPRT as well as determining the fractional male DNA concentration in the DNA mixtures. The dosage of *Zinc Finger* protein sequences on chromosome X (*ZFX*) and Y (*ZFY*) was determined by digital PCR analysis. An 87-bp amplicon of the *ZFX* and *ZFY* loci was first co-amplified by the forward primer 5'-CAAGTGCTGGACTCAGATGTAACTG-3' and the reverse primer 5'-TGAAGTAATGTCAGAAGCTAAAACATCA-3'. Two chromosome-specific TaqMan probes were designed to distinguish between the chromosome X and Y paralogs, and their sequences were 5'-(VIC)TCTTTAGCACATTGCA(MGBNFQ)-3' and 5'-(FAM)TCTTTACCACACTGCAC(MGBNFQ)-3', respectively.

[0263] The mutant dosage in the DNA mixtures was determined by digital PCR analysis of the normal allele relative to the mutant allele. For the HbE mutation, an 87-bp amplicon of

the normal and mutant alleles was first co-amplified by the forward primer 5'-GGGCAAGGTGAACGTGGAT -3' and the reverse primer 5'-CTATTGGTCTCCTTAAACCTGTCTTGTAA -3'. Two allele-specific TaqMan probes were designed to distinguish between the normal (G) and mutant (A) alleles, and their sequences were 5'-(VIC) TTGGTGGTGAGGCC (MGBNFQ)-3' and 5'-(FAM) TTGGTGGTAAGGCC (MGBNFQ)-3', respectively. Results for the HbE mutation are shown in FIG. 25.

[0264] For the CD41/42 deletion mutation, an 87- and 83-bp amplicon of the normal and mutant alleles was first co-amplified by the forward primer 5'-TTTTCCCACCCTTAGGCTGC -3' and the reverse primer 5'-ACAGCATCAGGAGTGGACAGATC -3', respectively. Two allele-specific TaqMan probes were designed to distinguish between the normal (without deletion) and mutant (with deletion) alleles, and their sequences were 5'-(VIC) CAGAGGTTCTTTGAGTCCT (MGBNFQ)-3' and 5'-(FAM) AGAGGTTGAGTCCTT (MGBNFQ)-3', respectively. Results for the HbE mutation are shown in FIGS. 26A and 26B.

[0265] These experiments were carried out on the BioMark™ System (Fluidigm) using the 12.765 Digital Arrays (Fluidigm). The reaction for one panel was set up using 2X TaqMan Universal PCR Master Mix (Applied Biosystems) in a reaction volume of 10 μL. For the CD41/42 and *ZFY/X* assays, each reaction contained 1X TaqMan Universal PCR Master Mix, 900 nM of each primer, 125 nM of each probe and 3.5 μL of DNA mixture at 1 ng/μL. For the HbE assay, 250 nM and 125 nM of probes targeting the normal (G) and mutant (A) alleles were added, respectively. The sample/ assay mixture was loaded into the Digital Array by the NanoFlex™ IFC controller (Fluidigm). The reaction was carried out on the BioMark™ System for signal detection. The reaction was initiated at 50°C for 2 min, followed by 95°C for 10 min and 50 cycles of 95°C for 15s and 57°C (for *ZFY/X* and CD41/42) or 56°C (for HbE) for 1 min. At least one reaction panel was used for each case, and data were aggregated from extra panels for samples which remained unclassified until a decision could be made.

[0266] It will also be obvious to those of skill in the art that the digital PCR can be performed using methods well-known to those of skill in the art, e.g. microfluidics chips, nanoliter PCR microplate systems, emulsion PCR, polony PCR, rolling-circle amplification, primer extension and mass spectrometry.

IX. EXAMPLE WITH CANCER

[0267] In one embodiment, the present invention may be performed to classify a sample as having allelic ratio skewing or not, as may occur in a cancerous tumor. In one aspect, for each case, the number of wells with positive signal for the A allele only, the G allele only, and both alleles were determined by digital PCR. The reference allele was defined as the allele with the smaller number of positive wells. (In the unlikely scenario that both alleles have the same number of positive wells, then either can be used as the reference allele.) The inferred average concentration of the reference allele per well (m_r) was calculated using the total number of wells negative for the reference allele, irrespective whether the other allele was positive, according to the Poisson probability density function. We use a hypothetical example to illustrate the calculation.

[0268] In a 96-well reaction, 20 wells are positive for the A allele, 24 wells are positive for the G allele, and 28 wells are positive for both alleles. The A allele would be regarded as the reference allele because less wells are positive for this allele. The number of wells negative for the reference allele would be 96 - 20 - 28 = 48. Therefore, m_r can be calculated using the Poisson distribution and would be -1n(48/96) = 0.693.

[0269] In the context of LOH detection, the null hypothesis refers to a sample that is assumed to lack allelic ratio skewing caused by the presence of a deletion of one allele. Under this assumption, the expected ratio of the number of positive wells for the two alleles would be 1:1 and, thus, the expected proportion of informative wells (wells positive for only one allele) containing the potentially overrepresented allele would be 0.5.

[0270] In the context of LOH detection, the alternative hypothesis refers to a sample that is assumed to have allelic ratio skewing caused by the presence of a deletion of one allele in 50% of the cells of the sample. As the allelic ratio between the overrepresented allele and the reference allele is 2:1, the average concentration of the overrepresented allele per well would be doubled that of the reference allele. However, the number of wells positive for the overrepresented allele would not be simply two times that for the reference allele but would follow the Poisson distribution.

[0271] An informative well is defined as a well positive for either the A or the G allele but not for both alleles. The calculation of the expected proportion of the number of wells containing the overrepresented alleles for samples with allelic ratio skewing is the same as is shown in Table 600. In the above example, if LOH is present in 50% of tumor cells, the

average concentration of the G allele per well would be 2 times 0.693 = 1.386. If LOH is present in more than 50% of the tumor cells, then the average concentration of the G allele per well would be according to the formula: 1/[1-(proportion with LOH)] × m_r.

[0272] The expected proportion of wells positive for the G allele would be 1 - $e^{-1..386}$= 0.75 (i.e., 75% or 72 wells). Assuming that the positivity of a well for the A or G allele is independent, 0.5 x 0.75 = 0.375 of the wells would be positive for both the A and G alleles. Hence, 0.5 – 0.375 = 0.125 of the wells would be positive for the A allele only and 0.75 – 0.375 = 0.375 of the wells would be positive for the G allele only. Therefore, the proportion of informative wells would be 0.125 + 0.375 = 0.5. The expected proportion of informative wells carrying the G allele would be 0.375/0.5 = 0.75. This expected value for P_r can then be used for the construction of appropriate SPRT curves for determining whether allelic ratio skewing (i.e. LOH in this context) is present in the sample.

[0273] The actual proportion of informative wells carrying the non-reference allele experimentallly determined by the digital PCR analysis (P_r) was then used to determine whether the null or alternative hypothesis would be accepted, or whether further analysis with more wells would be necessary. The decision boundaries for P_r to accept the null or alternative hypothesis was calculated based on a threshold likelihood ratio of 8 as this value had been shown to provide satisfactory performance to discriminate samples with and without allelic imbalance in the context of cancer detection (Zhou, W, et al. (2001) *Nat Biotechnol* 19, 78-81; Zhou et a1 2002, supra). In the above example, the number of informative wells would be 20 +24 = 44 and the experimentally obtained P_r would be 24/44 = 0.5455. The decision boundaries would be $\leq$0.5879 to accept the null hypothesis and $\geq$0.6739 to accept the alternative hypothesis. Therefore, the sample in this example would be classified as NOT having allelic ratio skewing.

[0274] In conclusion, we outlined an approach to detect sequence imbalance in a sample. In one embodiment, this invention can be used for the noninvasive detection of fetal chromosomal aneuploidy, such as trisomy 21 by analysis of fetal nucleic acids in maternal plasma. This approach can also be applied to other biological materials containing fetal nucleic acids, including amniotic fluid, chorionic villus samples, maternal urine, endocervical samples, maternal saliva, etc. First, we demonstrated the use of this invention for determining allelic imbalance of a SNP on *PLAC4* mRNA, a placenta-expressed transcript on chromosome 21, in maternal plasma of women bearing trisomy 21 fetuses. Second, we

demonstrated that our invention can be used as a non-polymorphism based method, through relative chromosome dosage (RCD) analysis, for the noninvasive prenatal detection of trisomy 21. Such a digital RCD-based approach involves the direct assessment of whether the total copy number of chromosome 21 in a sample containing fetal DNA is overrepresented with respect to a reference chromosome. Even without elaborate instrumentation, digital RCD allows the detection of trisomy 21 in samples containing 25% fetal DNA. We applied the sequential probability ratio test (SPRT) to interpret the digital PCR data. Computer simulation analyses confirmed the high accuracy of the disease classification algorithm.

[0275] We further outlined that the approach can be applied to the determination of other forms of nucleic acid sequence imbalances other than chromosomal aneuploidy, such as that for the detection of fetal mutation or polymorphism detection in maternal plasma and regional gains and losses in the genomes of malignant cells through the analysis of tumor-derived nucleic acids in plasma.

[0276] Any of the software components or functions described in this application, may be implemented as software code to be executed by a processor using any suitable computer language such as, for example, Java, C++ or Perl using, for example, conventional or object-oriented techniques. The software code may be stored as a series of instructions, or commands on a computer readable medium for storage and/or transmission, suitable media include random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard-drive or a floppy disk, or an optical medium such as a compact disk (CD) or DVD (digital versatile disk), flash memory, and the like. The computer readable medium may be any combination of such storage or transmission devices.

[0277] Such programs may also be encoded and transmitted using carrier signals adapted for transmission via wired, optical, and/or wireless networks conforming to a variety of protocols, including the Internet. As such, a computer readable medium according to an embodiment of the present invention may be created using a data signal encoded with such programs. Computer readable media encoded with the program code may be packaged with a compatible device or provided separately from other devices (e.g., via Internet download). Any such computer readable medium may reside on or within a single computer program product (e.g. a hard drive or an entire computer system), and may be present on or within different computer program products within a system or network. A computer system may

include a monitor, printer, or other suitable display for providing any of the results mentioned herein to a user.

[0278] An example of a computer system is shown in FIG. 27. The subsystems shown in FIG. 27 are interconnected via a system bus 2775. Additional subsystems such as a printer 2774, keyboard 2778, fixed disk 2779, monitor 2776, which is coupled to display adapter 2782, and others are shown. Peripherals and input/output (I/O) devices, which couple to I/O controller 2771, can be connected to the computer system by any number of means known in the art, such as serial port 2777. For example, serial port 2777 or external interface 2781 can be used to connect the computer apparatus to a wide area network such as the Internet, a mouse input device, or a scanner. The interconnection via system bus allows the central processor 2773 to communicate with each subsystem and to control the execution of instructions from system memory 2772 or the fixed disk 2779, as well as the exchange of information between subsystems. The system memory 2772 and/or the fixed disk 2779 may embody a computer readable medium.

[0279] The above description of exemplary embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form described, and many modifications and variations are possible in light of the teaching above. The embodiments were chosen and described in order to best explain the principles of the invention and its practical applications to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated.

[0280] All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

WHAT IS CLAIMED IS:

1. A method for determining whether a nucleic acid sequence imbalance exists within a biological sample, the method comprising:

receiving data from a plurality of reactions, wherein the data includes:

(1) a first set of quantitative data indicating a first amount of a clinically relevant nucleic acid sequence; and

(2) a second set of quantitative data indicating a second amount of a background nucleic acid sequence different from the clinically relevant nucleic acid sequence;

determining a parameter from the two data sets;

deriving a first cutoff value from an average concentration of a reference nucleic acid sequence in each of the plurality of reactions, wherein the reference nucleic acid sequence is either the clinically relevant nucleic acid sequence or the background nucleic acid sequence;

comparing the parameter to the first cutoff value; and

based on the comparison, determining a classification of whether a nucleic acid sequence imbalance exists.

2. The method of claim 1 wherein the first set of data are obtained from one or more first markers that each detect a presence of a part of the clinically relevant nucleic acid sequence in a reaction, and wherein the second set of data are obtained from one or more second markers that each detect a presence of a part of the background nucleic acid sequence in a reaction.

3. The method of claim 1, further comprising:

determining the average concentration of the reference nucleic acid sequence in each of the plurality of reaction using an inverse of a probability distribution having an input of a value derived from the data for the reference nucleic acid sequence.

4. The method of claim 1 wherein the clinically relevant nucleic acid sequence is from chromosome 21 and the background nucleic acid sequence is from a chromosome other than chromosome 21.

5. The method of claim 1 wherein the clinically relevant nucleic acid sequence is from chromosome 18 or 13 and the background nucleic acid sequence is from a chromosome other than chromosome 18 or 13 respectively.

6. The method of claim 1 wherein the clinically relevant nucleic acid sequence is an allele of a genetic polymorphism, and the background nucleic acid sequence is another allele of the genetic polymorphism.

7. The method of claim 1 wherein the clinically relevant nucleic acid sequence is a mutated copy of the cystic fibrosis transmembrane conductance regulator (CFTR) gene, the beta-globin gene or the alpha-globin gene, and the background nucleic acid sequence is from the wildtype copy of the corresponding gene.

8. The method of claim 1 wherein the biological sample is plasma or serum from a pregnant woman.

9. The method of claim 1 wherein a reaction is an amplification reaction.

10. The method of claim 9 wherein a reaction is a part of a digital PCR process.

11. The method of claim 1 wherein a reaction is a sequencing reaction.

12. The method of claim 1 wherein first portions of the clinically relevant nucleic acid sequence and the background nucleic acid sequence are from a first individual and second portions of the clinically relevant nucleic acid sequence and the background nucleic acid sequence are from a second individual.

13. The method of claim 12 wherein the cutoff value is based on a measurement of one of the first portions or a measurement of one of the second portions.

14. The method of claim 1, further comprising comparing the parameter to a second cutoff value.

15. The method of claim 14 wherein the classifications include disease state, non-disease state, and non-classifiable.

16. The method of claim 14 wherein the classifications include homozygous, heterozygous, and non-classifiable.

17. The method of claim 14 wherein the second cutoff value is based on a ratio of the first amount of the clinically relevant nucleic acid sequence relative to the second amount of the background nucleic acid sequence in a non-diseased state.

18. The method of claim 1 wherein the parameter is a ratio of the first amount of the clinically relevant nucleic acid sequence relative to the second amount of the background nucleic acid sequence.

19. The method of claim 1 wherein calculating the first cutoff value includes using at least one of: sequential probability ratio testing, false discovery rates, confidence intervals, and receiver operating characteristic curves.

20. The method of claim 1 wherein deriving the first cutoff value includes:

determining a proportion P1 of informative reactions containing an overrepresented nucleic acid sequence, which is either the reference or non-reference nucleic acid sequence; and

calculating the first cutoff value from the first proportion P1.

21. The method of claim 20 wherein determining the proportion P1 comprises:

determining a first probability of a reaction containing at least one of the overrepresented nucleic acid sequence;

calculating a second probability that a reaction is informative; and

calculating the proportion P1 using the first probability and the second probability.

22. The method of claim 21 wherein the first probability is determined by multiplying the average concentration of the reference nucleic acid sequence by an expected ratio relative to the non-reference nucleic acid sequence.

23. The method of claim 21 wherein the first probability is determined using the Poisson distribution having the average concentration of the overrepresented nucleic acid sequence in each of the plurality of reactions as an input.

24. The method of claim 21, further comprising:

determining a third probability of a reaction containing at least one of the underrepresented nucleic acid sequence, wherein calculating the second probability that the reaction is informative includes assuming that first probability and the second probability are independent.

25. A method for determining whether a nucleic acid sequence imbalance exists within a biological sample, the method comprising:

receiving data from a plurality of reactions, wherein the data includes:

(1) a first set of quantitative data indicating a first amount of a clinically relevant nucleic acid sequence; and

(2) a second set of quantitative data indicating a second amount of a background nucleic acid sequence different from the clinically relevant nucleic acid sequence, wherein the clinically relevant nucleic acid sequence and the background nucleic acid sequence come from a first type of cells and from one or more second types of cells;

determining a parameter from the two data sets;

deriving a first cutoff value from a first percentage resulting from a measurement of an amount of a nucleic acid sequence from the first type of cells in the biological sample;

comparing the parameter to the first cutoff value; and

based on the comparison, determining a classification of whether a nucleic acid sequence imbalance exists.

26. The method of claim 25 wherein the first type of cells is from a first organism and the second types of cells are from a second organism.

27. The method of claim 25 wherein deriving the first cutoff value includes:

determining a first average concentration of a reference nucleic acid sequence per reaction, wherein the reference nucleic acid sequence is the clinically relevant nucleic acid sequence or the background nucleic acid sequence that is underrepresented; and

multiplying the first average concentration by a factor derived from the first percentage to obtain a second average concentration of the nucleic acid sequence that is not the reference nucleic acid sequence.

28. The method of claim 27, further comprising:

determining the average concentration of the reference nucleic acid sequence in each of the plurality of reactions using an inverse of a probability distribution having an input of a value derived from the data for the reference nucleic acid sequence.

29. The method of claim 28 wherein the probability distribution is the Poisson distribution.

30. The method of claim 25 wherein the percentage is measured by determining an amount of a fetal-specific marker using quantitative real-time PCR, digital PCR, semiquantitative competitive PCR, real-competitive PCR, or mass spectrometry.

31. A computer program product comprising a computer readable medium encoded with a plurality of instructions for controlling a computing system to perform an operation for determining whether a nucleic acid sequence imbalance exists within a biological sample, the operation comprising the steps of:

receiving data from a plurality of reactions, wherein the data includes:

(1) a first set of quantitative data indicating a first amount of a clinically relevant nucleic acid sequence; and

(2) a second set of quantitative data indicating a second amount of a background nucleic acid sequence different from the clinically relevant nucleic acid sequence;

determining a parameter from the two data sets;

deriving a first cutoff value from an average concentration of a reference nucleic acid sequence in each of the plurality of reactions, wherein the reference nucleic acid sequence is either the clinically relevant nucleic acid sequence or the background nucleic acid sequence;

comparing the parameter to the first cutoff value; and

based on the comparison, determining a classification of whether a nucleic acid sequence imbalance exists.

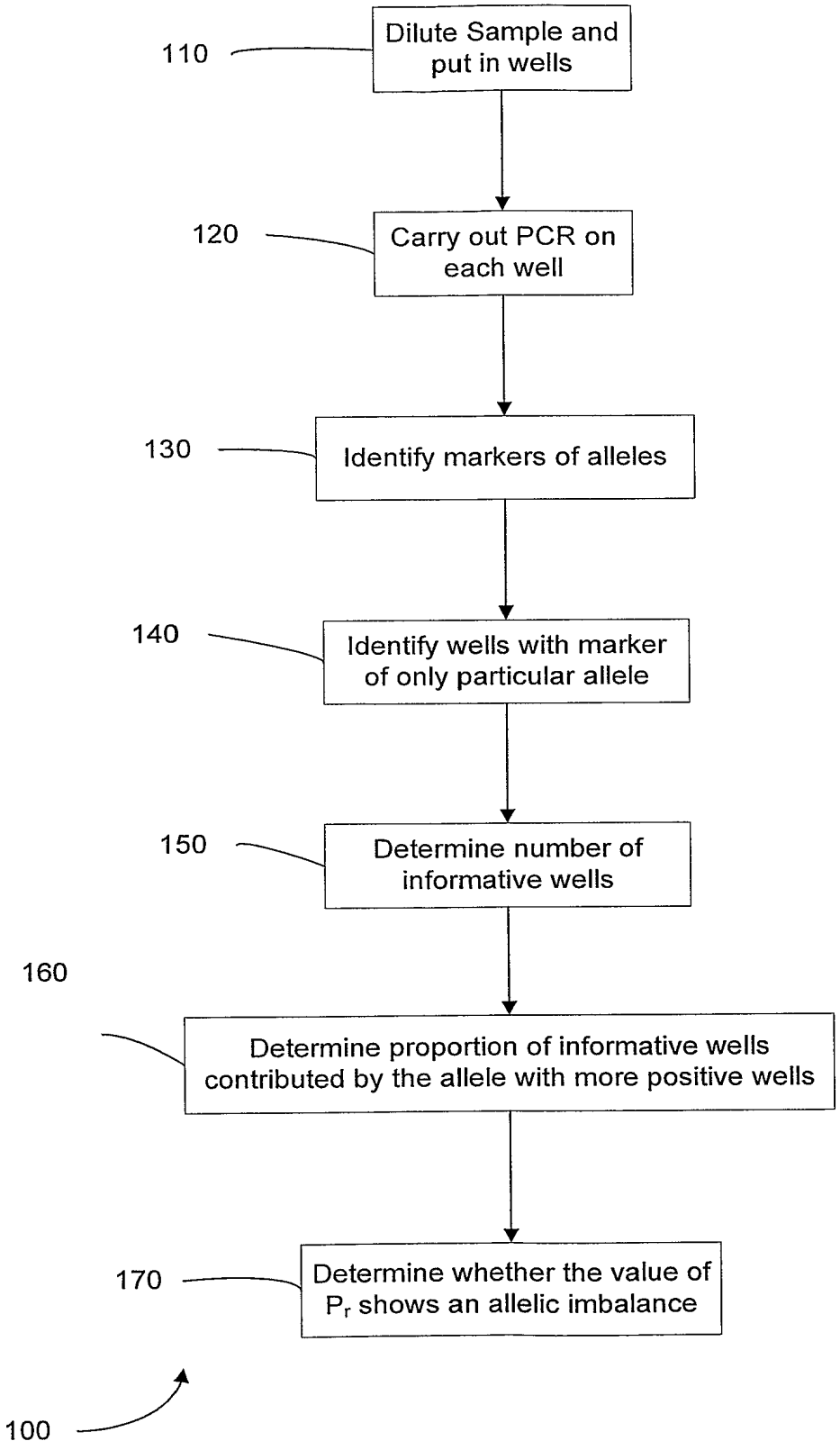

110
Dilute Sample and put in wells
120
Carry out PCR on each well
130
Identify markers of alleles
140
Identify wells with marker of only particular allele
150
Determine number of informative wells
160
Determine proportion of informative wells contributed by the allele with more positive wells
170
Determine whether the value of P_r shows an allelic imbalance
100

FIG. 1

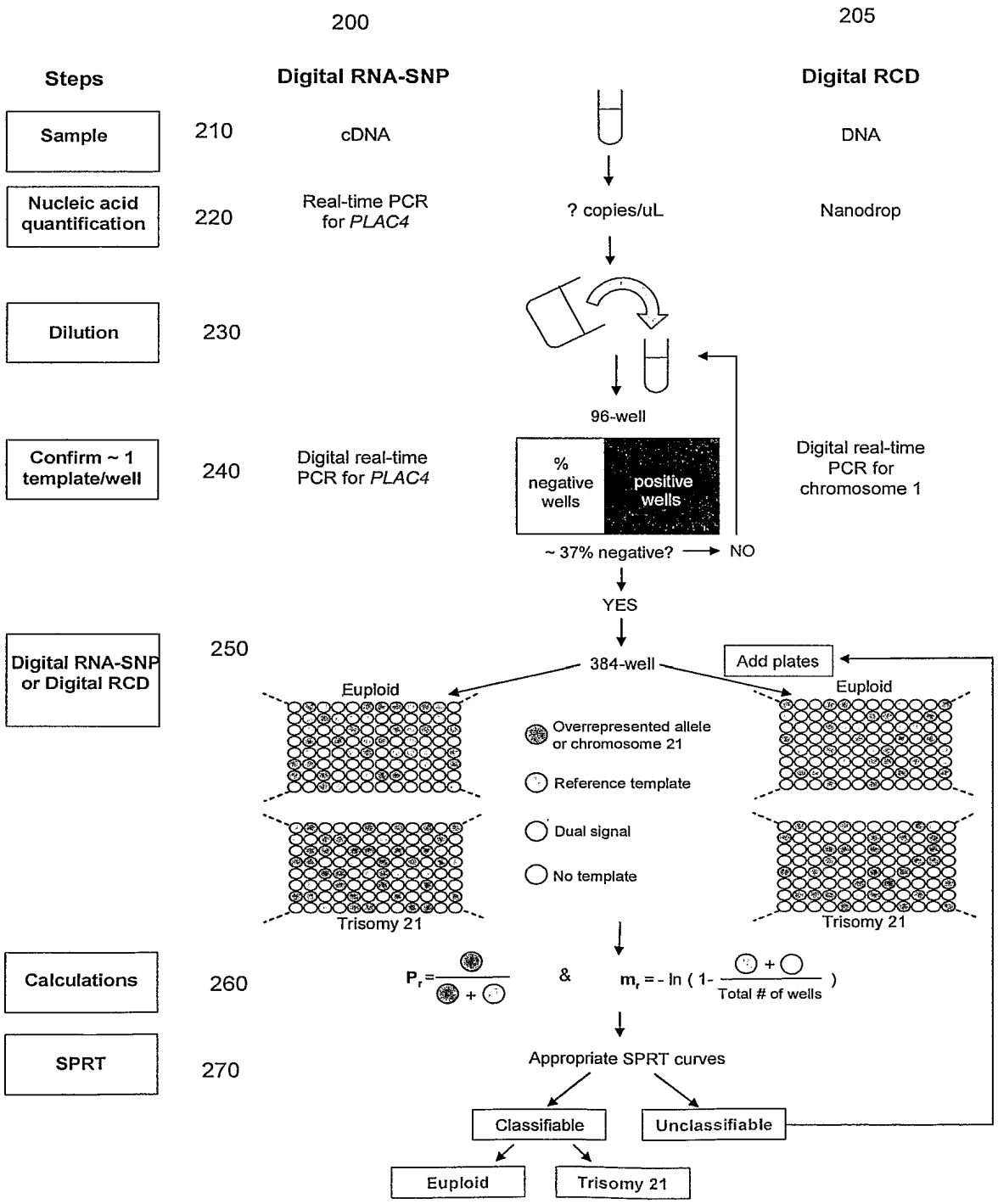

200
205
Steps
Digital RNA-SNP
Digital RCD
Sample
210
cDNA
DNA
Nucleic acid quantification
220
Real-time PCR for PLAC4
? copies/uL
Nanodrop
Dilution
230
96-well
Confirm ~ 1 template/well
240
Digital real-time PCR for PLAC4
% negative wells
positive wells
Digital real-time PCR for chromosome 1
~ 37% negative?
NO
YES
Digital RNA-SNP or Digital RCD
250
384-well
Add plates
Euploid
Euploid
Overrepresented allele or chromosome 21
Reference template
Dual signal
No template
Trisomy 21
Trisomy 21
Calculations
260
$P_r = \frac{\text{Overrepresented}}{\text{Overrepresented} + \text{Reference}}$
&
$m_r = -\ln\left(1 - \frac{\text{Reference} + \text{Dual}}{\text{Total \# of wells}}\right)$
SPRT
270
Appropriate SPRT curves
Classifiable
Unclassifiable
Euploid
Trisomy 21

FIG. 2A

Cancer	Common chromosomal aberrations	
	Gain	Loss
Bladder cancer	7	9p; 11q; 17p
Breast cancer	1q; 8q; 17q; 20q	8p; 13q; 16q; 18q
Colorectal cancer	7; 8q; 13q; 20q	1p; 5q; 8p; 15q; 17p; 18q
Hepatocellular carcinoma	1q; 8q; 20q	4q; 8p; 13q; 16q; 17p
Non - small cell lung carcinoma	1q; 3q; 5p; 8q	3p; 8p; 9p; 13q; 17p
Small cell lung cancinoma	3q; 5p; 8q; 19q	3p; 4q; 5q; 10q; 13q; 16q
Nasopharyngeal carcinoma	1q; 3q; 8; 12; 19	1p; 3p; 9; 11q; 13q; 14q; 16q
Prostate cancer	7; 8q; Xq	8p; 13q; 16q

FIG. 2B

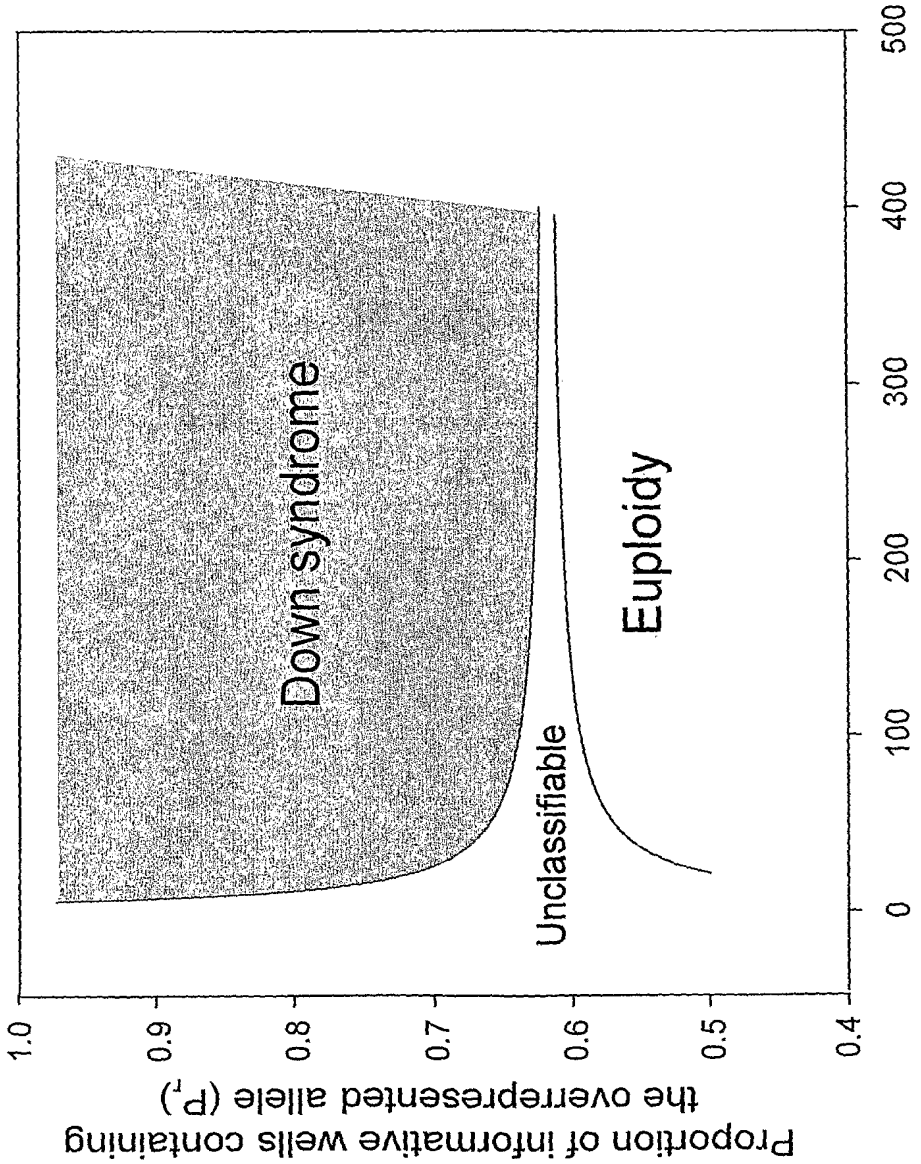

Down syndrome
Unclassifiable
Euploidy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
100
200
300
400
500
No. of informative wells
Proportion of informative wells containing the overrepresented allele (P_r)

FIG. 3

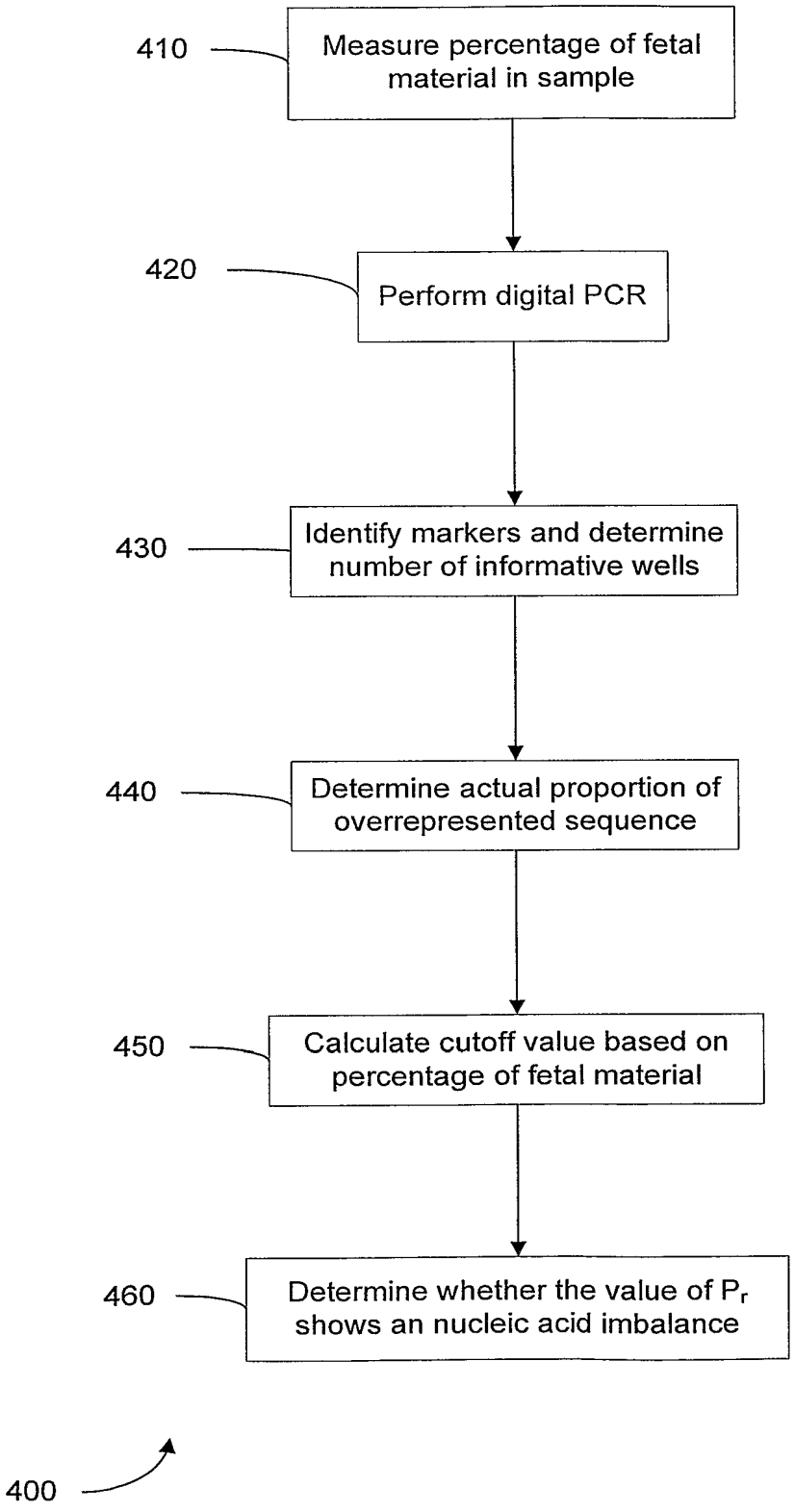

410
Measure percentage of fetal material in sample
420
Perform digital PCR
430
Identify markers and determine number of informative wells
440
Determine actual proportion of overrepresented sequence
450
Calculate cutoff value based on percentage of fetal material
460
Determine whether the value of P_r shows an nucleic acid imbalance
400

FIG. 4

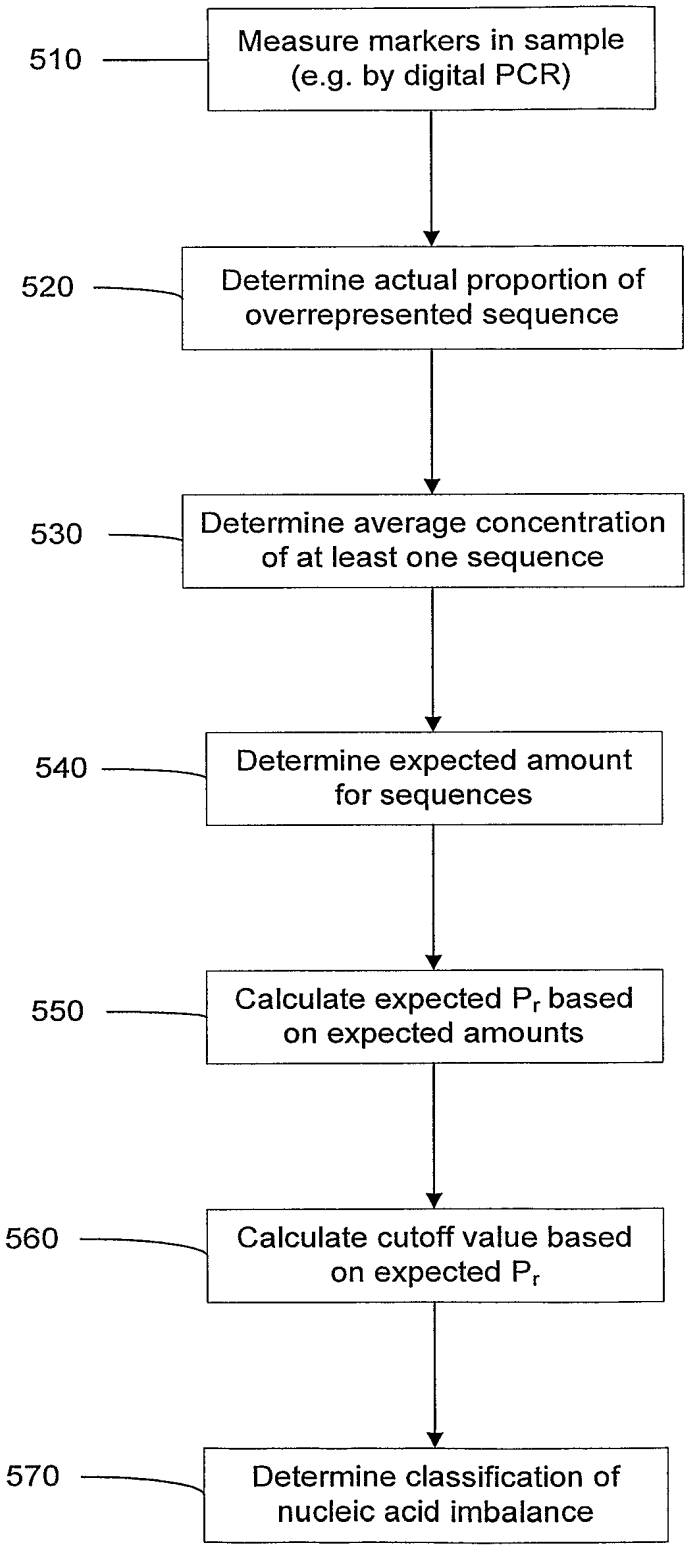

510
Measure markers in sample (e.g. by digital PCR)
520
Determine actual proportion of overrepresented sequence
530
Determine average concentration of at least one sequence
540
Determine expected amount for sequences
550
Calculate expected P_r based on expected amounts
560
Calculate cutoff value based on expected P_r
570
Determine classification of nucleic acid imbalance
500

FIG. 5

m_r	Proportion of wells positive for the reference allele	Proportion of wells positive for the overrepresented allele	Proportion of wells positive for both alleles	Proportion of wells positive for the reference allele ONLY	Proportion of wells positive for the overrepresented allele ONLY	Digital RNA-SNP allelic ratio	Proportion of informative wells	P_r
0.1	0.0952	0.1813	0.0173	0.0779	0.164	2.11	0.2419	0.68
0.2	0.1813	0.3297	0.0598	0.1215	0.2699	2.22	0.3914	0.69
0.3	0.2592	0.4512	0.1169	0.1422	0.3342	2.35	0.4765	0.7
0.4	0.3297	0.5507	0.1815	0.1481	0.3691	2.49	0.5173	0.71
0.5	0.3935	0.6321	0.2487	0.1448	0.3834	2.65	0.5282	0.73
0.6	0.4512	0.6988	0.3153	0.1359	0.3835	2.82	0.5194	0.74
0.7	0.5034	0.7534	0.3793	0.1241	0.3741	3.01	0.4983	0.75
0.8	0.5507	0.7981	0.4395	0.1112	0.3586	3.23	0.4698	0.76
0.9	0.5934	0.8347	0.4953	0.0981	0.3394	3.46	0.4375	0.78
1	0.6321	0.8647	0.5466	0.0855	0.3181	3.72	0.4036	0.79
1.1	0.6671	0.8892	0.5932	0.0739	0.296	4	0.3699	0.8
1.2	0.6988	0.9093	0.6354	0.0634	0.2739	4.32	0.3373	0.81
1.3	0.7275	0.9257	0.6734	0.054	0.2523	4.67	0.3063	0.82
1.4	0.7534	0.9392	0.7076	0.0458	0.2316	5.06	0.2774	0.83
1.5	0.7769	0.9502	0.7382	0.0387	0.212	5.48	0.2507	0.85
1.6	0.7981	0.9592	0.7656	0.0325	0.1937	5.95	0.2262	0.86
1.7	0.8173	0.9666	0.79	0.0273	0.1766	6.47	0.2039	0.87
1.8	0.8347	0.9727	0.8119	0.0228	0.1608	7.05	0.1836	0.88
1.9	0.8504	0.9776	0.8314	0.019	0.1462	7.69	0.1652	0.88
2	0.8647	0.9817	0.8488	0.0158	0.1329	8.39	0.1487	0.89

600

FIG. 6

m_f	Percentage of fetal DNA in the sample							
	10%		25%		50%		100%	
	Percentage of informative wells	P_r	Percentage of informative wells	P_r	Percentage of informative wells	P_r	Percentage of informative wells	P_r
0.1	18%	0.51	18%	0.53	19%	0.56	21%	0.61
0.2	30%	0.51	31%	0.53	32%	0.56	35%	0.61
0.3	39%	0.51	40%	0.53	41%	0.57	43%	0.62
0.4	45%	0.51	45%	0.54	46%	0.57	48%	0.63
0.5	48%	0.52	49%	0.54	49%	0.57	51%	0.63
0.6	50%	0.52	50%	0.54	50%	0.58	51%	0.64
0.7	50%	0.52	50%	0.54	50%	0.58	50%	0.65
0.8	49%	0.52	49%	0.54	49%	0.58	48%	0.65
0.9	48%	0.52	47%	0.55	47%	0.59	46%	0.66
1	46%	0.52	45%	0.55	44%	0.59	43%	0.67
1.1	44%	0.52	43%	0.55	42%	0.6	40%	0.68
1.2	41%	0.52	40%	0.55	39%	0.6	37%	0.69
1.3	39%	0.52	38%	0.55	36%	0.6	34%	0.69
1.4	36%	0.52	35%	0.56	33%	0.61	31%	0.7
1.5	34%	0.52	33%	0.56	31%	0.61	28%	0.71
1.6	31%	0.52	30%	0.56	28%	0.62	26%	0.72
1.7	29%	0.53	28%	0.56	26%	0.62	23%	0.73
1.8	27%	0.53	25%	0.57	24%	0.63	21%	0.73
1.9	24%	0.53	23%	0.57	21%	0.63	19%	0.74
2	22%	0.53	21%	0.57	20%	0.64	17%	0.75

700

FIG. 7

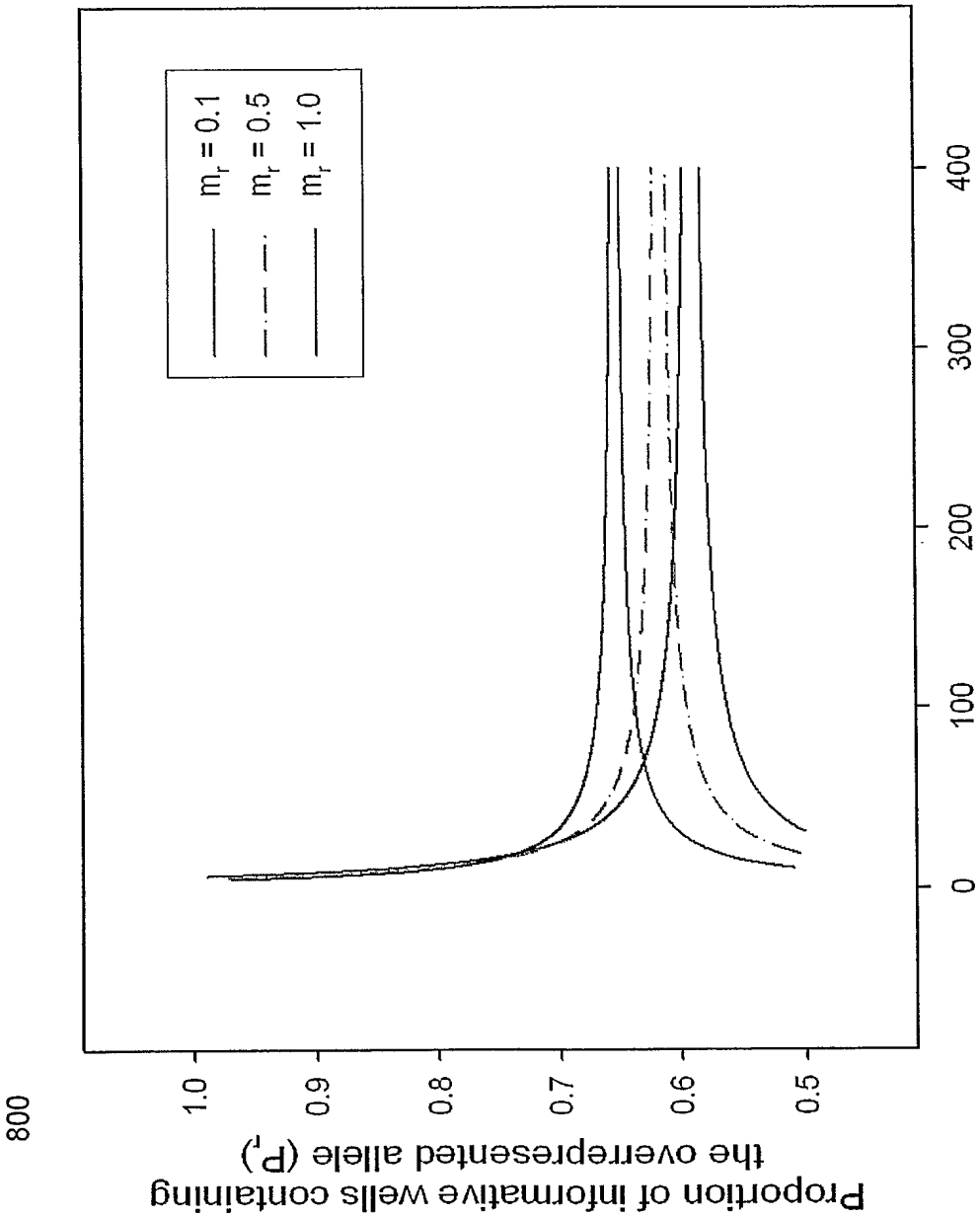

Proportion of informative wells containing the overrepresented allele (P_I)
1.0
0.9
0.8
0.7
0.6
0.5
0
100
200
300
400
No. of informative wells
m_r = 0.1
m_r = 0.5
m_r = 1.0
800

FIG. 8

m_r	Classification	New algorithm		Old algorithm		m_r	Classification	New algorithm		Old algorithm	
		Euploid	Trisomy 21	Euploid	Trisomy 21			Euploid	Trisomy 21	Euploid	Trisomy 21
	Correct	0%	37%	0%	35%		correct	89%	92%	14%	35%
	Incorrect	3%	2%	3%	1%		Incorrect	1%	0%	4%	0%
0.1	Unclassified	97%	61%	97%	63%	1.1	Unclassified	10%	7%	81%	65%
	Correct	6%	64%	0%	40%		correct	90%	92%	11%	35%
	Incorrect	4%	2%	4%	0%		Incorrect	1%	0%	4%	0%
0.2	Unclassified	90%	34%	96%	60%	1.2	Unclassified	9%	8%	84%	65%
	Correct	34%	75%	6%	39%		correct	91%	92%	7%	34%
	Incorrect	4%	2%	4%	0%		Incorrect	1%	1%	4%	0%
0.3	Unclassified	62%	24%	90%	61%	1.3	Unclassified	8%	7%	89%	66%
	Correct	54%	81%	14%	39%		correct	92%	91%	5%	34%
	Incorrect	3%	1%	4%	0%		Incorrect	1%	1%	4%	0%
0.4	Unclassified	43%	19%	82%	61%	1.4	Unclassified	8%	9%	91%	66%
	Correct	64%	85%	20%	38%		correct	92%	90%	2%	32%
	Incorrect	3%	1%	4%	0%		Incorrect	1%	1%	4%	0%
0.5	Unclassified	33%	14%	76%	62%	1.5	Unclassified	8%	9%	93%	68%
	Correct	74%	88%	23%	38%		correct	92%	89%	1%	31%
	Incorrect	2%	1%	4%	0%		Incorrect	1%	1%	4%	0%
0.6	Unclassified	24%	11%	72%	62%	1.6	Unclassified	8%	10%	94%	69%
	Correct	80%	89%	25%	37%		correct	92%	89%	1%	30%
	Incorrect	2%	1%	4%	0%		Incorrect	1%	1%	4%	0%
0.7	Unclassified	18%	10%	71%	63%	1.7	Unclassified	7%	10%	96%	70%
	Correct	83%	91%	23%	37%		correct	92%	88%	0%	28%
	Incorrect	2%	1%	4%	0%		Incorrect	1%	1%	4%	0%
0.8	Unclassified	15%	9%	73%	63%	1.8	Unclassified	7%	11%	96%	72%
	Correct	86%	91%	21%	36%		correct	92%	87%	0%	26%
	Incorrect	1%	1%	4%	0%		Incorrect	1%	1%	3%	0%
0.9	Unclassified	12%	8%	75%	64%	1.9	Unclassified	8%	12%	97%	74%
	Correct	88%	92%	19%	36%		correct	91%	86%	0%	24%
	Incorrect	1%	0%	4%	0%		Incorrect	1%	1%	4%	0%
1	Unclassified	11%	8%	77%	64%	2	Unclassified	9%	13%	96%	76%

900

FIG. 9A

m_r	Classification	New algorithm		Old algorithm		m_r	Classification	New algorithm		Old algorithm	
		Euploid	Trisomy 21	Euploid	Trisomy 21			Euploid	Trisomy 21	Euploid	Trisomy 21
	correct	57%	88%	48%	89%		correct	100%	100%	92%	90%
	Incorrect	3%	1%	3%	1%		Incorrect	0%	0%	1%	0%
0.1	Unclassified	40%	11%	49%	10%	1.1	Unclassified	0%	0%	7%	10%
	correct	89%	98%	79%	90%		correct	100%	100%	90%	90%
	Incorrect	1%	0%	2%	0%		Incorrect	0%	0%	1%	0%
0.2	Unclassified	10%	2%	19%	10%	1.2	Unclassified	0%	0%	9%	10%
	correct	96%	99%	88%	90%		correct	100%	100%	89%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	1%	0%
0.3	Unclassified	4%	1%	11%	10%	1.3	Unclassified	0%	0%	10%	10%
	correct	99%	100%	92%	90%		correct	100%	100%	87%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	1%	0%
0.4	Unclassified	1%	0%	7%	10%	1.4	Unclassified	0%	0%	12%	10%
	correct	100%	100%	94%	90%		correct	100%	100%	85%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	1%	0%
0.5	Unclassified	0%	0%	6%	10%	1.5	Unclassified	0%	0%	13%	10%
	correct	100%	100%	94%	90%		correct	100%	100%	82%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	1%	0%
0.6	Unclassified	0%	0%	5%	10%	1.6	Unclassified	0%	0%	17%	10%
	correct	100%	100%	95%	90%		correct	100%	100%	79%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	2%	0%
0.7	Unclassified	0%	0%	5%	10%	1.7	Unclassified	0%	0%	19%	10%
	correct	100%	100%	95%	90%		correct	100%	100%	74%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	2%	0%
0.8	Unclassified	0%	0%	5%	10%	1.8	Unclassified	0%	0%	24%	10%
	correct	100%	100%	94%	90%		correct	100%	100%	70%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	2%	0%
0.9	Unclassified	0%	0%	6%	10%	1.9	Unclassified	0%	0%	28%	10%
	correct	100%	100%	94%	90%		correct	100%	100%	67%	90%
	Incorrect	0%	0%	1%	0%		Incorrect	0%	0%	3%	0%
1	Unclassified	0%	0%	6%	10%	2	Unclassified	0%	0%	30%	10%

950

FIG. 9B

m_r	Classification result	*Euploid*	*T21*	m_r	Classification result	*Euploid*	*T21*
0.1	Correct	57%	88%	0.9	Correct	100%	100%
	Incorrect	3%	1%		Incorrect	0%	0%
	Unclassified	40%	11%		Unclassified	0%	0%
0.2	Correct	89%	98%	1	Correct	100%	100%
	Incorrect	1%	0%		Incorrect	0%	0%
	Unclassified	10%	2%		Unclassified	0%	0%
0.3	Correct	96%	99%	1.1	Correct	100%	100%
	Incorrect	0%	0%		Incorrect	0%	0%
	Unclassified	4%	1%		Unclassified	0%	0%
0.4	Correct	99%	100%	1.2	Correct	100%	100%
	Incorrect	0%	0%		Incorrect	0%	0%
	Unclassified	1%	0%		Unclassified	0%	0%
0.5	Correct	100%	100%	1.3	Correct	100%	100%
	Incorrect	0%	0%		Incorrect	0%	0%
	Unclassified	0%	0%		Unclassified	0%	0%
0.6	Correct	100%	100%	1.4	Correct	100%	100%
	Incorrect	0%	0%		Incorrect	0%	0%
	Unclassified	0%	0%		Unclassified	0%	0%
0.7	Correct	100%	100%	1.5	Correct	100%	100%
	Incorrect	0%	0%		Incorrect	0%	0%
	Unclassified	0%	0%		Unclassified	0%	0%
0.8	Correct	100%	100%	2	Correct	100%	100%
	Incorrect	0%	0%		Incorrect	0%	0%
	Unclassified	0%	0%		Unclassified	0%	0%

1000

FIG. 10

m_r	Classification result	*Euploid*	*Trisomy 21*	m_r	Classification result	*Euploid*	*Trisomy 21*
	Correct	38%	44%		Correct	95%	94%
	Incorrect	2%	2%		Incorrect	0%	1%
0.1	Unclassified	60%	54%	**0.7**	Unclassified	5%	5%
	Correct	64%	69%		Correct	95%	95%
	Incorrect	2%	2%		Incorrect	1%	0%
0.2	Unclassified	34%	29%	**0.8**	Unclassified	4%	5%
	Correct	78%	82%		Correct	96%	96%
	Incorrect	2%	1%		Incorrect	0%	0%
0.3	Unclassified	20%	17%	**0.9**	Unclassified	4%	4%
	Correct	85%	87%		Correct	97%	96%
	Incorrect	1%	1%		Incorrect	0%	0%
0.4	Unclassified	14%	12%	**1**	Unclassified	3%	4%
	Correct	89%	90%		Correct	97%	96%
	Incorrect	1%	1%		Incorrect	0%	0%
0.5	Unclassified	10%	9%	**1.5**	Unclassified	3%	4%
	Correct	92%	93%		Correct	96%	95%
	Incorrect	1%	1%		Incorrect	0%	0%
0.6	Unclassified	7%	7%	**2**	Unclassified	4%	5%

1100

FIG. 11

Number of wells	Classification results	Fetal DNA percentage in the sample							
		10%		25%		50%		100%	
		Euploid	*Trisomy 21*	*Euploid*	*Trisomy 21*	*Euploid*	*Trisomy 21*	*Euploid*	*Trisomy 21*
384	Correct	0%	0%	6%	7%	47%	49%	89%	90%
	Incorrect	0%	0%	0%	1%	2%	2%	1%	1%
	Unclassified	100%	100%	93%	92%	51%	49%	10%	9%
768	Correct	0%	0%	23%	24%	73%	76%	98%	99%
	Incorrect	0%	0%	2%	2%	2%	1%	0%	0%
	Unclassified	100%	100%	75%	74%	25%	23%	2%	1%
1152	Correct	1%	1%	39%	40%	86%	87%	100%	100%
	Incorrect	0%	0%	2%	2%	1%	1%	0%	0%
	Unclassified	99%	99%	59%	58%	13%	12%	0%	0%
1536	Correct	2%	2%	50%	51%	93%	93%	100%	100%
	Incorrect	0%	0%	2%	2%	1%	1%	0%	0%
	Unclassified	98%	97%	48%	47%	7%	6%	0%	0%
1920	Correct	4%	4%	60%	60%	96%	96%	100%	100%
	Incorrect	0%	1%	2%	2%	0%	0%	0%	0%
	Unclassified	96%	95%	38%	37%	4%	3%	0%	0%
3840	Correct	18%	19%	84%	83%	100%	100%	100%	100%
	Incorrect	1%	1%	1%	1%	0%	0%	0%	0%
	Unclassified	81%	80%	15%	15%	18%	22%	0%	0%
7680	Correct	43%	45%	97%	97%	100%	100%	100%	100%
	Incorrect	2%	2%	0%	0%	0%	0%	0%	0%
	Unclassified	55%	53%	3%	3%	0%	0%	0%	0%

1200

FIG. 12

sample	geno-type[a]	no. of wells	no. of wells positive for individual alleles				m_r[b]	SPRT result		
			A only	G only	AG	all negative		P_r[c]	unclassifiable region[d]	classification[e]
Placental DNA										
N677	AG	384	85	83	126	90	0.79	0.51	0.63 - 0.65	euploid
N710	AG	384	102	83	73	126	0.52	0.55	0.61 - 0.63	euploid
N435	AGG	384	49	157	130	48	0.63	0.76	0.62 - 0.64	T21
N981	AAG	384	135	69	82	98	0.5	0.66	0.61 - 0.63	T21
Placental RNA										
V533	AG	384	103	93	71	117	0.56	0.53	0.61 - 0.63	euploid
V943	AG	384	89	100	74	121	0.55	0.53	0.61 - 0.63	euploid
N435	AGG	384	52	138	95	99	0.48	0.73	0.61 - 0.63	T21
T215	AAG	384	146	58	138	42	0.71	0.72	0.62 - 0.64	T21

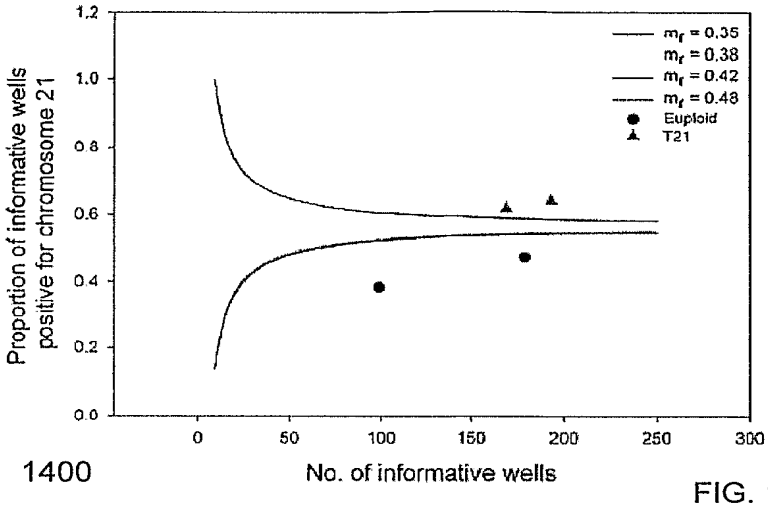
1300

FIG. 13A

sample	geno-type	no. of wells	no. of wells positive for individual alleles				m	SPRT result		
			A only	G only	AG	all negative		ratio	unclassifiable region	classification
M2390P	AG	384	90	100	97	97	0.67	0.526	0.62 - 0.64	euploid
M2391P	AG	384	97	105	65	117	0.55	0.52	0.61 - 0.63	euploid
M2473P	AG	384	66	92	34	192	0.3	0.582	0.59 - 0.62	euploid
M2524P	AG	384	29	28	3	324	0.08	0.509	0.54 - 0.64	euploid
M2528P	AG	384	112	85	44	143	0.41	0.569	0.60 - 0.62	euploid
M2601P	AG	384	90	101	72	121	0.55	0.529	0.61 - 0.63	euploid
M2607P	AG	384	73	91	57	163	0.41	0.555	0.60 - 0.63	euploid
M2638P	AG	384	66	90	52	176	0.37	0.577	0.59 - 0.62	euploid
M2639P	AG	384	71	56	17	240	0.21	0.559	0.58 - 0.62	euploid
M2525P	AAG	384	110	53	21	200	0.21	0.675	0.58 - 0.61	T21
M2272P	AAG	768	246	127	112	283	0.37	0.66	0.60 - 0.61	T21
M2718P	AGG	384	66	114	66	138	0.42	0.633	0.60 - 0.62	T21
M1519P	AGG	384	58	130	54	142	0.34	0.691	0.59 - 0.62	T21

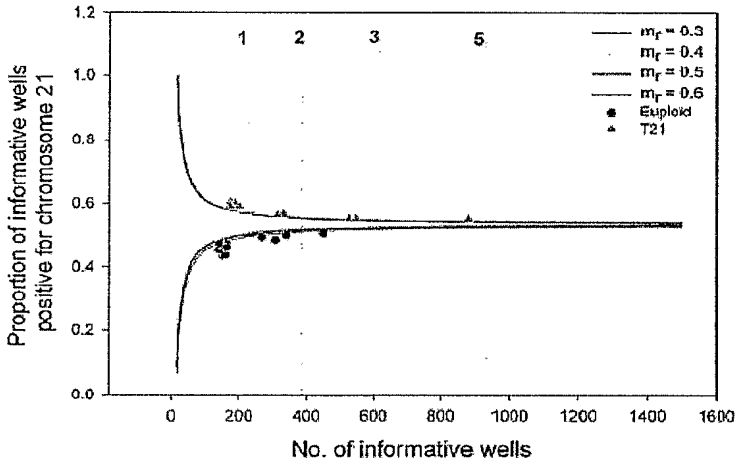
1350

FIG. 13B



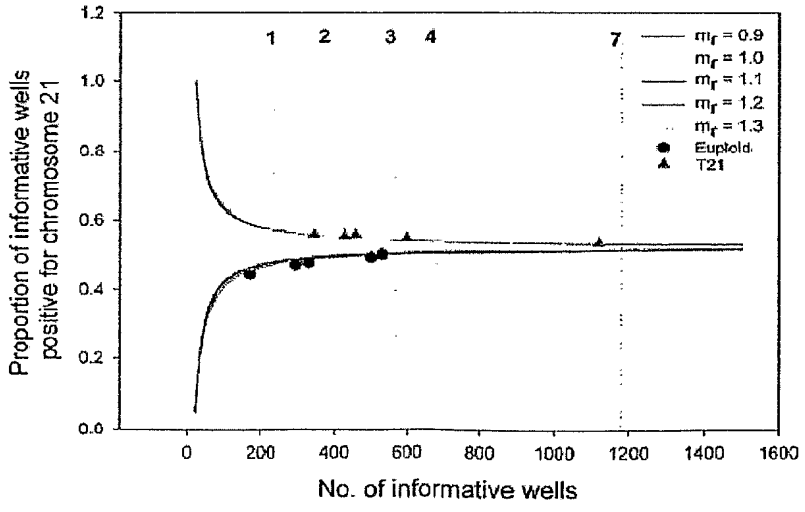
1400

FIG. 14A



1430

FIG. 14B



1460

FIG. 14C

sample	genotype	no. of wells	no. of wells positive for individual alleles				m_r[a]	P_r[b]	SPRT result	
			A only	G only	GA	all negative			unclassifiable region	classification[c]
V943	AG	765	168	209	117	271	0.47	0.554	0.609- 0.621	euploid
N435	AGG	765	92	281	105	287	0.30	0.753	0.597- 0.610	T21
T21-5	AAG	765	267	126	141	231	0.43	0.679	0.607- 0.618	T21

[a] no. of reference molecules per well

[b] P_r values were calculated by the equation:

$$\frac{\text{no. of wells positive for the overrepresented allele}}{\text{no. of wells positive for A only + no of wells positive for G only}}$$

[c] "Euploid" was assigned when the Pr was below the unclassifiable region; "T21" was assigned when the Pr was abov e the unclassifiable region. T21, trisomy 21

FIG. 15A

panel #	no. of wells	no. wells positive for individual alleles			
		A only	G only	GA	all negative
S1	765	14	15	0	736
S2	765	7	10	0	748
S3	765	13	9	1	744
S4	765	7	8	0	750
S5	765	12	8	0	745
S6	765	8	14	0	743
S7	765	13	9	0	743
S8	765	9	8	0	748
S9	765	14	15	0	736
S10	765	9	12	0	744
S11	765	8	17	0	740
S12	765	11	9	0	745

FIG. 15B

sample	geno-type	panels[a]	no. of wells positive for individual alleles					SPRT result		
			no of wells	A only	G only	GA	all negative	P_r	unclassifiable region	classification
M2879P	AG	5	3825	82	103	5	3635	0.557	0.570- 0.602	Euploid
M2875P	AG	3	2295	35	40	0	2220	0.533	0.546- 0.625	Euploid
M3078P	AG	3	2295	33	33	1	2228	0.500	0.541- 0.631	Euploid
M2976P	AG	12	9180	56	46	0	9078	0.549	0.556- 0.615	Euploid
M2831P	AAG	2	1530	42	13	0	1475	0.764	0.531- 0.640	T21

[a] no. of panels needed for classification

FIG. 15C

Sample	Genotype[a]	No. of wells positive for A or G allele			m_r[b]	SPRT analysis		
		A only	G only	Both		P_r[c]	Unclassifiable region[d]	Classification[e]
N3505	AG	157	131	48	0.27	0.545	0.59-0.61	Euploid
N3530	AG	174	132	52	0.28	0.569	0.59-0.61	Euploid
N3531	AG	146	129	28	0.23	0.531	0.58-0.60	Euploid
N3533	AG	149	135	46	0.27	0.525	0.59-0.61	Euploid
N2772	AG	159	149	79	0.35	0.516	0.60-0.61	Euploid
N3506	AG	154	145	66	0.32	0.515	0.60-0.61	Euploid
N3555	AG	94	73	7	0.11	0.563	0.58-0.61	Euploid
N3560	AG	101	109	28	0.18	0.519	0.58-0.61	Euploid
N3561	AG	91	70	17	0.12	0.565	0.58-0.61	Euploid
N828	AAG	153	89	22	0.16	0.632	0.58-0.60	T18
CA2	AGG	12	28	0	0.016	0.700	0.51-0.66	T18
CA7	AGG	97	171	56	0.22	0.638	0.59-0.61	T18

[a] Genotypes were determined by mass spectrometry.

[b] The m_r value indicates the average number of reference template molecules per well.

[c] The P_r value indicates the proportion of informative wells positive for the overrepresented allele.

[d] The unclassifiable region varies with the m_r value.

[e] "Euploid" was assigned when P_r was below the unclassifiable region; "T18" was assigned when P_r was above the unclassifiable region.

FIG. 16A

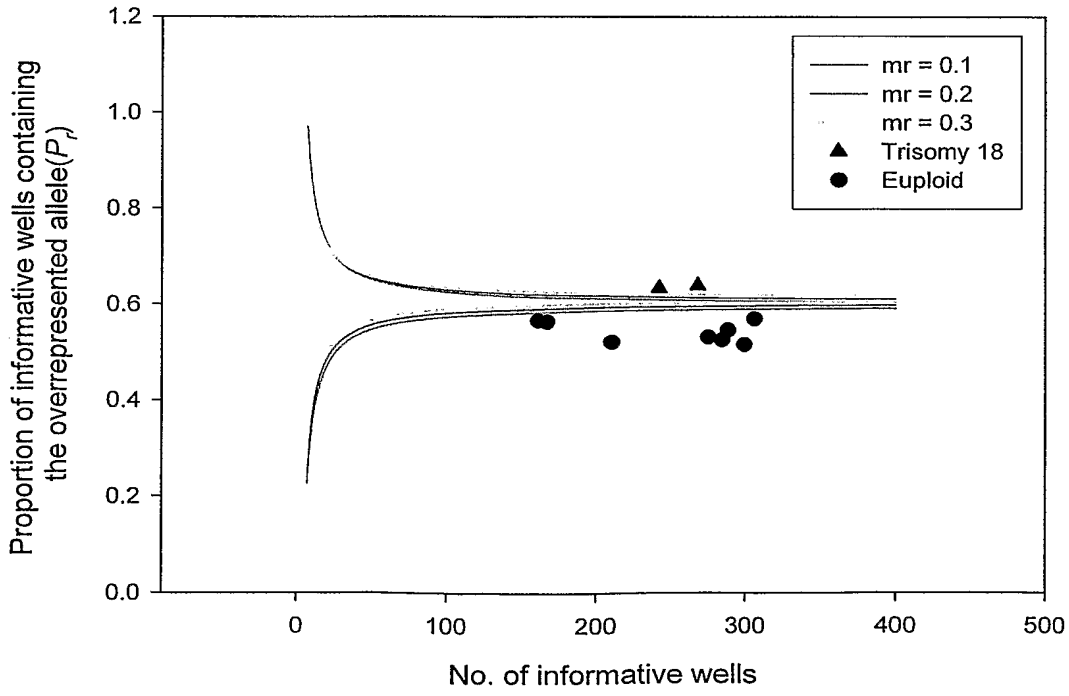


FIG. 16B

Type	Cases	Total No. of panels used	Total No. of informative wells	No. of well count			m_r[a]	SPRT analysis		
				Both positive	Chr21 only	Chr1 only		P_r[b]	Unclassifiable region	Classification[c]
T21	N0234	1	303	68	172	131	0.30	0.568	0.507 - 0.559	T21
	N0274	2	558	118	307	251	0.28	0.550	0.518 - 0.547	T21
	N0332	1	293	73	168	125	0.30	0.573	0.506 - 0.560	T21
	N0349	2	574	107	322	252	0.27	0.561	0.518 - 0.546	T21
	N0435	1	299	67	178	123	0.29	0.589	0.506 - 0.559	T21
	N0456	1	318	68	189	129	0.30	0.594	0.508 - 0.558	T21
	N0667	2	602	134	345	257	0.30	0.573	0.519 - 0.546	T21
	N0764	1	289	62	167	122	0.28	0.578	0.505 - 0.560	T21
	N0890	4	1206	272	657	549	0.31	0.545	0.526 - 0.539	T21
	N0891	1	317	103	180	137	0.38	0.568	0.510 - 0.558	T21
Normal	N0196	1	262	68	111	151	0.34	0.424	0.504 - 0.563	Euploid
	N0198	1	295	68	132	163	0.34	0.463	0.506 - 0.561	Euploid
	N0230	2	627	109	272	255	0.27	0.516	0.517 - 0.548	Euploid
	N0231	2	542	98	275	267	0.27	0.507	0.517 - 0.547	Euploid
	N0262	1	295	57	144	151	0.32	0.488	0.506 - 0.560	Euploid
	N0463	1	263	70	120	143	0.33	0.456	0.504 - 0.563	Euploid
	N0527	1	292	56	137	155	0.32	0.469	0.506 - 0.560	Euploid
	N0619	1	281	60	138	143	0.31	0.491	0.505 - 0.561	Euploid
	N0625	1	291	60	134	157	0.33	0.460	0.507 - 0.560	Euploid
	N0651	2	566	118	287	279	0.30	0.507	0.519 - 0.547	Euploid

[a] no. of reference molecules per well

[b] P_r values were calcualated by the equation:

$$\frac{\text{no. of wells positive for the overrepresented allele (chr21)}}{\text{no. of wells positive for chr21 only + no of wells positive for chr1 only}}$$

[c] "Euploid" was assigned when the Pr was below the unclassifiable region; "T21" was assigned when the Pr was above the unclassifiable region.

FIG. 17

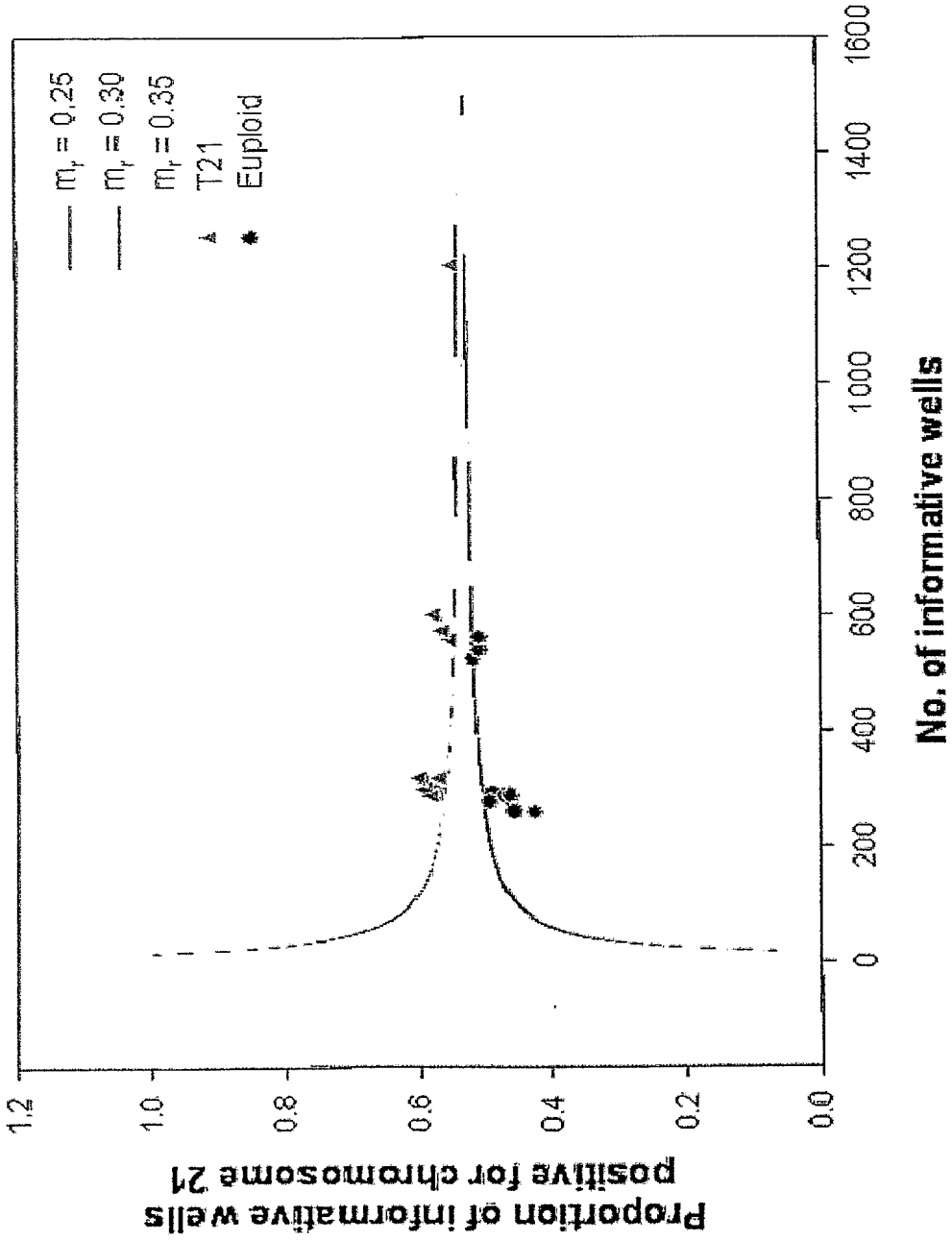

Proportion of informative wells positive for chromosome 21
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
1400
1600
No. of informative wells
m_r = 0.25
m_r = 0.30
m_r = 0.35
T21
Euploid

FIG. 18

Type	Cases	No. of panels used	No. of informative wells	No. of well count			m_r[a]	SPRT analysis		
				Both positive	Chr21 only	Chr1 only		P_r[b]	Unclassifiable region	Classification[c]
T21	A0274	1	380	269	249	131	0.74	0.655	0.567 - 0.585	T21
	A2008	1	348	373	233	115	1.02	0.670	0.579 - 0.596	T21
	A2178	1	384	288	239	145	0.83	0.622	0.571 - 0.588	T21
	AS363A	1	364	109	244	120	0.36	0.670	0.551 - 0.573	T21
	AS979A	1	256	492	188	88	1.32	0.734	0.590 - 0.610	T21
	AS3457A	1	376	83	256	120	0.31	0.681	0.549 - 0.572	T21
Normal	A0236	1	387	275	195	192	0.94	0.504	0.576 - 0.592	Euploid
	A0237	1	392	228	184	208	0.84	0.469	0.572 - 0.588	Euploid
	A0255	1	378	222	195	183	0.75	0.516	0.568 - 0.585	Euploid
	A0269	1	254	49	127	127	0.25	0.500	0.541 - 0.578	Euploid
	AS365A	1	370	135	192	178	0.53	0.519	0.558 - 0.578	Euploid
	AS483A	1	354	84	188	188	0.38	0.525	0.550 - 0.574	Euploid
	AS484A	1	372	271	194	178	0.88	0.522	0.573 - 0.590	Euploid
	AS780A	1	413	188	200	213	0.74	0.484	0.568 - 0.584	Euploid
	AS780A	1	316	67	159	157	0.35	0.503	0.548 - 0.575	Euploid
	AS897A	1	198	8	87	111	0.17	0.439	0.532 - 0.579	Euploid
	AS921A	1	152	8	89	83	0.13	0.454	0.523 - 0.586	Euploid
	AS980A	1	394	137	191	203	0.59	0.485	0.561 - 0.580	Euploid
	AS2282A	1	334	342	177	157	1.06	0.530	0.580 - 0.597	Euploid
	AS2425A	1	128	8	64	64	0.10	0.500	0.516 - 0.591	Euploid
	AS2434A	1	379	97	203	176	0.44	0.536	0.555 - 0.575	Euploid
	A1119	1	354	123	185	169	0.48	0.523	0.555 - 0.577	Euploid
	A1122	1	369	100	195	174	0.44	0.528	0.554 - 0.576	Euploid
	A1152	1	256	48	128	128	0.26	0.500	0.542 - 0.578	Euploid
	A1180	1	246	498	133	113	1.61	0.541	0.604 - 0.622	Euploid
	A1275	1	373	95	201	172	0.43	0.539	0.554 - 0.575	Euploid
	A1375	1	359	55	190	169	0.35	0.529	0.550 - 0.573	Euploid
	A1388	1	357	87	165	182	0.47	0.462	0.555 - 0.577	Euploid
	A1399	1	371	114	197	174	0.47	0.531	0.556 - 0.577	Euploid

[a] no. of reference molecules per well

[b] P_r values were calcualated by the equation:

$$\frac{\text{no. of wells positive for the overrepresented allele (chr21)}}{\text{no. of wells positive for chr21 only + no of wells positive for chr1 only}}$$

[c] "Euploid" was assigned when the Pr was below the unclassifiable region; "T21" was assigned when the Pr was above the unclassifiable region.

FIG. 19

Type	Cases	Total No. of panels used	Total No. of informative wells	No. of well count			m_r [a]	SPRT analysis		
				Both positive	Chr18 only	Chr21 only		P_r [b]	Unclassifiable region	Classification [c]
Euploid	N0230	1	380	243	181	199	0.86	0.476	0 .573 - 0.589	E
	N0625	1	370	225	179	191	0.78	0.484	0 .569 - 0.586	E
	N0651	1	358	206	192	166	0.67	0.536	0.564 - 0.583	E
T18	N0828	1	543	256	329	214	0.95	0.606	0.579 - 0.590	T18
	N0830	1	346	353	240	106	0.92	0.694	0.574 - 0.592	T18
	N0936	1	372	285	209	163	0.88	0.562	0.573 - 0.590	E
	N1678	1	343	115	221	122	0.37	0.644	0.551 - 0.574	T18
	N1701	1	303	94	191	112	0.31	0.630	0.546 - 0.575	T18

[a] no. of reference molecules per well

[b] P_r values were calcualated by the equation:

$$\frac{\text{no. of wells positive for the overrepresented allele (chr18)}}{\text{no. of wells positive for chr21 only + no of wells positive for chr18 only}}$$

[c] "Euploid" was assigned when the Pr was below the unclassifiable region; "T18" was assigned when the Pr was above the unclassifiable region.

FIG. 20

Type	Cases	Plate No.	Assay	Total No. of informative wells	No. of well count			m_r[a]	SPRT analysis		
					Both positive	Chr 21 only	Ref only		P_r[b]	Unclassifiable region	Classification[c]
Normal	N0196	1	chr21/chr1	148	78	71	77	0.52	0.480	0.489 - 0.584	E
		1	chr21/chr18	152	63	79	73	0.44	0.520	0.487 - 0.584	U
			SUM	300	141	150	150	0.48	0.500	0.512 - 0.560	E
	N0198	1	chr21/chr1	162	73	70	92	0.56	0.432	0.490 - 0.581	E
		1	chr21/chr18	183	68	93	90	0.53	0.508	0.498 - 0.575	U
			SUM	345	141	163	182	0.55	0.472	0.517 - 0.557	E
	N0230	1	chr21/chr1	162	58	81	81	0.45	0.500	0.490 - 0.581	U
		1	chr21/chr18	165	45	81	84	0.41	0.491	0.489 - 0.580	U
			SUM	327	103	162	165	0.43	0.495	0.512 - 0.558	E
	N0231	1	chr21/chr1	173	53	81	92	0.47	0.468	0.494 - 0.578	E
		1	chr21/chr18	166	51	85	81	0.42	0.512	0.490 - 0.580	U
			SUM	339	104	166	173	0.45	0.490	0.514 - 0.557	E
	N0262	1	chr21/chr1	186	97	101	85	0.64	0.543	0.503 - 0.575	U
		1	chr21/chr18	176	111	91	85	0.71	0.517	0.504 - 0.557	U
			SUM	362	208	192	170	0.68	0.530	0.522 - 0.557	U
		2	chr21/chr1	174	103	85	89	0.69	0.489	0.503 - 0.577	E
			SUM	536	311	277	259	0.68	0.517	0 .523 - 0.552	E
T21	N0234	1	chr21/chr1	193	68	108	85	0.51	0.560	0.500 - 0.573	U
		1	chr21/chr18	178	65	92	86	0.50	0.517	0.496 - 0.576	U
			SUM	371	133	200	171	0.50	0.539	0.517 - 0.556	U
		2	chr21/chr1	178	73	92	86	0.53	0.517	0.498 - 0.576	U
			SUM	549	206	292	257	0.51	0.532	0 .524 - 0.549	U
		2	chr21/chr18	178	62	95	83	0.47	0.534	0.495 - 0.576	U
			SUM	727	268	387	340	0.50	0.532	0 .527 - 0.546	U
		3	chr21/chr1	184	108	103	81	0.68	0.560	0.504 - 0.575	U
			SUM	911	376	490	421	0.54	0.538	0.529 - 0.545	U
		3	chr21/chr18	198	102	120	78	0.63	0.606	0.505 - 0.572	T21
			SUM	1109	478	610	499	0.55	0.550	0.531 - 0.544	T21
	N0332	1	chr21/chr1	173	109	103	70	0.63	0.595	0.500 - 0.577	T21
		1	chr21/chr18	191	82	110	81	0.55	0.576	0.501 - 0.574	T21
			SUM	364	191	213	151	0.59	0.585	0.519 - 0.557	T21
	N0349	1	chr21/chr1	158	62	83	75	0.44	0.525	0.489 - 0.582	U
		1	chr21/chr18	166	46	87	79	0.39	0.524	0.489 - 0.580	U
			SUM	324	108	170	154	0.42	0.525	0.512 - 0.558	U
		2	chr21/chr1	149	66	86	63	0.41	0.577	0.485 - 0.585	U
			SUM	473	174	256	217	0.41	0.541	0.519 - 0.551	U
		2	chr21/chr18	168	43	100	68	0.34	0.595	0.487 - 0.580	T21
			SUM	641	217	356	285	0.40	0.555	0.523 - 0.546	T21
	N0764	1	chr21/chr1	182	82	106	76	0.53	0.582	0.498 - 0.575	T21
		1	chr21/chr18	178	83	97	81	0.56	0.545	0.498 - 0.576	U
			SUM	360	165	203	157	0.54	0.564	0.518 - 0.556	T21
	N0890	1	chr21/chr1	180	103	108	72	0.61	0.600	0.501 - 0.576	T21
		1	chr21/chr18	188	68	104	84	0.50	0.553	0.499 - 0.574	U
			SUM	368	171	212	156	0.55	0.576	0.518 - 0.556	T21

[a] no. of reference molecules per well

[b] P_r values were calcualated by the equation:

$$\frac{\text{no. of wells positive for the overrepresented allele (chr21)}}{\text{no. of wells positive for chr21 only + no of wells positive for Ref only}}$$

[c] "Euploid" was assigned when the Pr was below the unclassifiable region; "T21" was assigned when the Pr was above the unclassifiable region.

FIG. 21

			Poisson corrected total no. Chromosome 21	Reference chr	P_r	Classification
Euploid	Case 1	Assay 1	37	33	0.52345751	**Unclass**
		Assay 2	31	42	0.42506276	**Unclass**
		Assay 3	29	37	0.44198785	**Unclass**
		Assay 4	39	26	0.6008691	**Unclass**
		Plate 1 subtotal	136	138	0.495821	**Unclass**
		Assay 1	32	25	0.56605316	**Unclass**
		Assay 2	27	24	0.53165424	**Unclass**
		Assay 3	10	17	0.3681535	**Unclass**
		Assay 4	24	18	0.5626764	**Unclass**
		Plate 2 subtotal	93	84	0.52474	**Unclass**
		Total	229	223	0.507182	**Euploid**
	Case 2	Assay 1	45	35	0.56321585	**Unclass**
		Assay 2	31	45	0.41245525	**Unclass**
		Assay 3	32	22	0.59954763	**Unclass**
		Assay 4	38	35	0.5228175	**Unclass**
		Plate 1 subtotal	146	135	0.519153	**Unclass**
		Assay 1	28	23	0.55272135	**Unclass**
		Assay 2	23	22	0.51196923	**Unclass**
		Assay 3	23	15	0.59688187	**Unclass**
		Assay 4	29	26	0.5293349	**Unclass**
		Plate 2 subtotal	102	85	0.545204	**Unclass**
		Plate 1+2 subtotal	248	221	0.529583	**Unclass**
		Assay 1	17	25	0.4122149	**Unclass**
		Assay 2	25	16	0. 6027299	**Unclass**
		Assay 3	13	26	0.3382802	**Unclass**
		Assay 4	27	24	0.5316872	**Unclass**
		Plate 3 subtotal	82	91	0.475775	**Unclass**
		Plate 1- 3 subtotal	331	311	0.515069	**Unclass**
		Assay 1	15	9	0.6269193	**Unclass**
		Assay 2	14	15	0.4824155	**Unclass**
		Assay 3	18	17	0.5146313	**Unclass**
		Assay 4	21	18	0.5270080	**Unclass**
		Plate 4 subtotal	69	60	0.532264	**Unclass**
		Plate 1- 4 subtotal	399	372	0.517942	**Unclass**

FIG. 22A

	Assay 1	10	13	0.4337957	**Unclass**
	Assay 2	11	14	0.4390082	**Unclass**
	Assay 3	7	16	0.3018022	**Unclass**
	Assay 4	19	10	0.6579085	**Unclass**
	Plate 5 subtotal	48	54	0.469951	**Unclass**
	Total	447	426	0.5123444	**Euploid**
T21					
Case 3	Assay 1	50	33	0.60012253	**Unclass**
	Assay 2	56	47	0.54458055	**Unclass**
	Assay 3	46	28	0.6196626	**Unclass**
	Assay 4	56	36	0.6108731	**Unclass**
	Total	208	144	0.59079	**T21**
Case 4	Assay 1	56	36	0.61085513	**Unclass**
	Assay 2	38	32	0.53906912	**Unclass**
	Assay 3	40	24	0.62788068	**Unclass**
	Assay 4	47	27	0.634511	**Unclass**
	Total	180	119	0.603476	**T21**

FIG. 22B

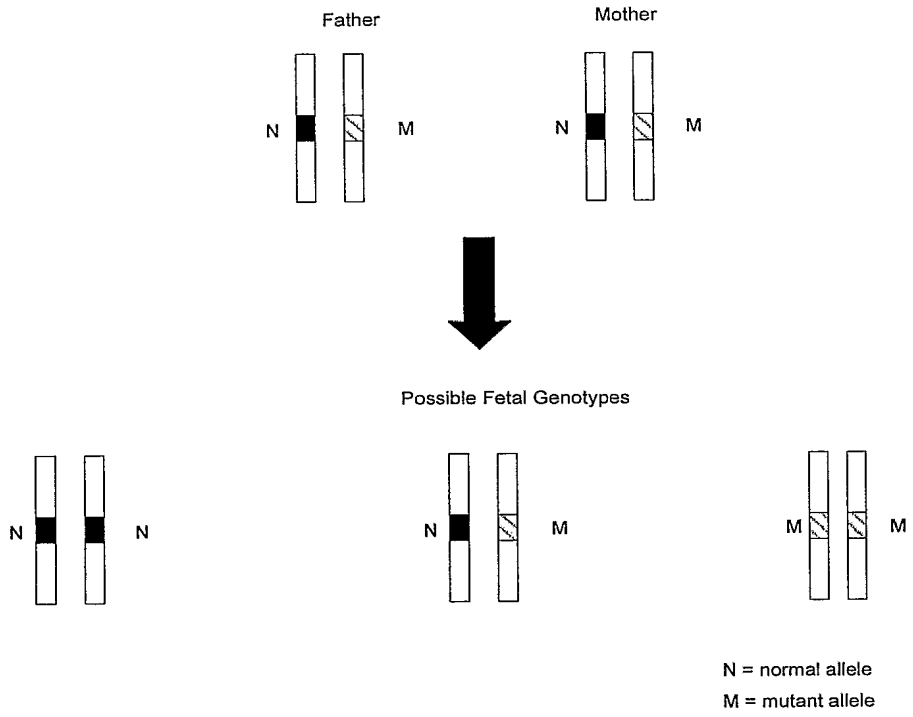

Father
Mother
N
M
N
M
Possible Fetal Genotypes
N
N
N
M
M
M
N = normal allele
M = mutant allele

FIG. 23

Cases	Samples	Expected %	No of panels	*ZFY*	*ZFX*	ALL POS	NET *ZFY*	NET *ZFX*	Ref allele	m_r	SPRT[a]
1	XY + XX	50%	1	239	465	155	84	310	Y	0.37	XX
	(XX %)	25%	1	328	447	205	123	242	Y	0.56	XX
2	$XY_1 + XY_2$	50%	1	414	405	220	194	185	X	0.85	XY
	(XY_2 %)	25%	1	432	426	245	187	181	X	0.81	XY

FIG. 24A

Cases	No of panels	*ZFY*	*ZFX*	ALL POS	NET *ZFY*	NET *ZFX*	Ref allele	m_r	SPRT
1	1	81	171	18	63	153	Y	0.11	XX
2	1	103	175	23	80	152	Y	0.14	XX
3	1	110	175	28	82	147	Y	0.16	XX
4	1	111	156	24	87	132	Y	0.16	XX
5	2	236	322	50	186	272	Y	0.17	XX
6	1	116	184	29	87	155	Y	0.16	XX
7	1	94	169	19	75	150	Y	0.13	XX
8	1	91	155	24	67	131	Y	0.13	XX
9	3	333	447	74	259	373	Y	0.16	XX
10	1	110	154	24	86	130	Y	0.16	XX
11	1	83	170	26	57	144	Y	0.11	XX
12	1	85	159	16	69	143	Y	0.12	XX

FIG. 24B

Cases	No of panels	Maternal genotype[a]	Fetal genotype[a]	Expected fetal %	A	G	ALL POS	NET A	NET G	Actual fetal %[b]	Ref allele	m_t	SPRT[d]
1	1	AG	GG	15%	264	318	116	148	202	9.6	A	0.42	GG
2	1	AG	GG	15%	208	276	81	127	195	14.3	A	0.32	GG
3	1	AG	AG	15%	248	251	87	161	164	15.4	A	0.39	AG
4	1	AG	GG	15%	158	240	51	107	189	13.7	A	0.23	GG
5	1	AG	AG	15%	226	245	70	156	175	16.8	A	0.35	AG
6	1	AG	AG	15%	319	319	130	189	189	18.1	A	0.54	AG
7	1	AG	AG	15%	254	279	94	160	185	19.8	A	0.40	AG
8	1	AG	GG	15%	296	363	144	152	219	15.7	A	0.49	GG
9	1	AG	AG	15%	279	285	100	179	185	13.0	A	0.45	AG
10	1	AG	AA	15%	337	278	120	217	158	NA[c]	G	0.45	AA
1	1	AG	GG	20%	248	295	96	152	199	13.0	A	0.39	GG
2	1	AG	GG	20%	191	289	79	112	210	19.5	A	0.29	GG
3	1	AG	AG	20%	248	218	75	173	143	18.2	G	0.34	U
4	1	AG	GG	20%	158	257	59	99	198	25.9	A	0.23	GG
5	1	AG	AG	20%	231	215	71	160	144	20.3	G	0.33	AG
6	1	AG	AG	20%	281	283	98	183	185	26.3	A	0.46	AG
7	1	AG	AG	20%	256	231	88	168	143	22.7	G	0.36	AG
8	1	AG	GG	20%	246	311	104	142	207	23.0	A	0.39	GG
9	1	AG	AG	20%	267	249	83	184	166	21.9	G	0.39	AG
10	1	AG	AA	20%	303	203	77	226	126	NA[c]	G	0.31	AA
1	1	AG	GG	25%	244	304	99	145	205	15.7	A	0.38	GG
2	1	AG	GG	25%	209	314	88	121	226	24.9	A	0.32	GG
3	1	AG	AG	25%	204	212	55	149	157	28.0	A	0.31	AG
4	1	AG	GG	25%	185	290	76	109	214	25.9	A	0.28	GG
5	1	AG	AG	25%	233	243	77	156	166	24.5	A	0.36	AG
6	1	AG	AG	25%	292	305	110	182	195	27.3	A	0.48	AG
7	1	AG	AG	25%	261	275	95	166	180	22.5	A	0.42	AG
8	1	AG	GG	25%	236	327	96	140	231	26.7	A	0.37	GG
9	1	AG	AG	25%	271	280	104	167	176	25.2	A	0.44	AG
10	1	AG	AA	25%	243	154	46	197	108	NA[c]	G	0.22	AA
1	1	AG	GG	50%	136	293	57	79	236	33.6	A	0.20	GG
2	1	AG	GG	50%	163	358	78	85	280	45.8	A	0.24	GG
3	1	AG	AG	50%	200	208	53	147	155	55.7	A	0.30	AG
4	1	AG	GG	50%	137	356	66	71	290	53.0	A	0.18	GG
5	1	AG	AG	50%	247	246	91	156	155	57.7	G	0.39	AG
6	1	AG	AG	50%	218	245	68	150	177	53.8	A	0.34	AG
7	1	AG	AG	50%	255	282	103	152	179	51.3	A	0.41	AG
8	1	AG	GG	50%	158	398	83	75	315	57.6	A	0.23	GG
9	1	AG	AG	50%	235	225	74	161	151	55.7	G	0.35	AG
10	1	AG	AA	50%	317	106	35	282	71	NA[c]	G	0.15	AA

[a] A = Mutated allele, G = Wildtype allele.
[b] The actual fetal DNA % is determined by the digital *ZFYIX* assay.
[c] The fetus is female and digital *ZFYIX* assay cannot be applied.
[d] SPRT was done with the fetal % determined by the digital *ZFYIX* assay except case 10; U, unclassified.

FIG. 25

Case	Maternal genotype[a]	Fetal genotype[a]	No of panels	M	N	ALL POS	NET M	NET N	1[b] Expected fetal %	1[b] SPRT	2[c] Fetal % by digital *ZFY/X*	2[c] SPRT	Ref Allele	m_r
1	MN	MM	3	846	763	273	573	490	5%	MM	7.8%	MM	N	0.40
	MN	MM	3	863	748	289	574	459	10%	MM	9.6%	MM	N	0.39
	MN	MM	1	323	219	79	244	140	25%	MM	24.1%	MM	N	0.34
	MN	MM	1	418	160	92	326	68	50%	MM	57.2%	MM	N	0.23
2	MN	MN	9	2559	2476	887	1672	1589	5%	U	3.4%	U	N	0.45
	MN	MN	3	895	907	360	535	547	10%	MN	9.7%	MN	M	0.49
	MN	MN	1	277	263	97	180	166	25%	MN	24.2%	MN	N	0.42
	MN	MN	1	273	295	101	172	194	50%	MN	58.0%	MN	M	0.44

[a] M = with CTTT deletion, N = without CTTT deletion.

[b] SPRT was done with the expected %.

[c] SPRT was done with the fetal % determined by the digital *ZFY/X* assay; U, unclassified

FIG. 26A

Cases	No of panels	Maternal genotype[a]	Fetal genotype[a]	Expected fetal %	M	N	ALL POS	NET M	NET N	Actual fetal %[b]	Ref allele	m_r	SPRT[c]
1	1	MN	MN	20%	203	200	58	145	142	29.0	N	0.30	MN
2	1	MN	NN	20%	149	204	39	110	165	29.7	M	0.22	NN
3	1	MN	NN	20%	128	212	31	97	181	30.7	M	0.18	NN
4	1	MN	MN	20%	173	196	45	128	151	27.0	M	0.26	MN
5	1	MN	NN	20%	147	230	47	100	183	24.3	M	0.21	NN
6	1	MN	MN	20%	199	206	45	154	161	19.2	M	0.30	MN
7	1	MN	NN	20%	165	237	58	107	179	18.7	M	0.24	NN
8	1	MN	MN	20%	203	180	43	160	137	23.4	N	0.27	MN
9	1	MN	NN	20%	173	241	61	112	180	16.9	M	0.26	NN
10	1	MN	MN	20%	174	190	52	122	138	16.8	M	0.26	U
11	1	MN	NN	20%	167	242	49	118	193	17.7	M	0.25	NN
12	1	MN	MN	20%	213	199	47	166	152	21.6	N	0.30	MN

[a] M = with CTTT deletion, N = without CTTT deletion.

[b] The actual fetal DNA % is determined by the digital *ZFXY* assay.

[c] SPRT was done with the fetal % determined by the digital *ZFY/X* assay; U, unclassified.

FIG. 26B

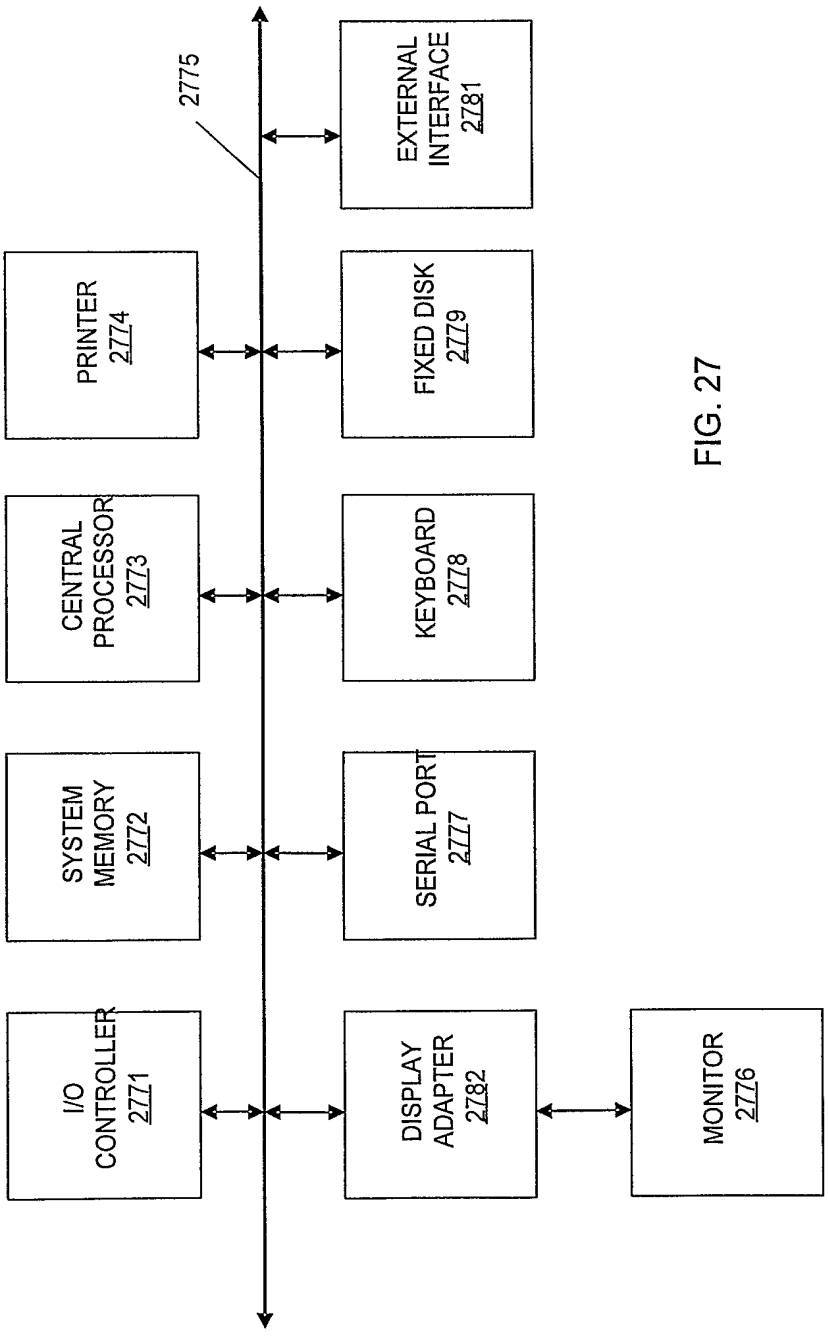

2775
EXTERNAL INTERFACE 2781
PRINTER 2774
FIXED DISK 2779
CENTRAL PROCESSOR 2773
KEYBOARD 2778
SYSTEM MEMORY 2772
SERIAL PORT 2777
I/O CONTROLLER 2771
DISPLAY ADAPTER 2782
MONITOR 2776

FIG. 27

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2008/002524

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/00 C12Q1/68

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)
G06F C12Q

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)
EPO-Internal, BIOSIS, EMBASE, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHOU W ET AL: "Counting alleles to predict recurrence of early-stage colorectal cancers" LANCET THE, LANCET LIMITED. LONDON, GB, vol. 359, no. 9302, 19 January 2002 (2002-01-19), pages 219-225, XP004791874 ISSN: 0140-6736 cited in the application abstract page 220, right-hand column, line 38 - page 221, right-hand column, line 10 figures 1,2 ----- -/--	1-12, 14-18, 25,26, 30,31

[X] 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 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
24 October 2008	30/10/2008

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, Fax: (+31-70) 340-3016	Kürten, Ivayla

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHIH I-M ET AL: "Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis" CANCER RESEARCH, vol. 61, February 2002 (2002-02), pages 818-822, XP007905911 page 819, right-hand column, line 13 - page 820, right-hand column, line 21 figure 2	1-12, 14-18, 25,26, 30,31
X	TONG YU K ET AL: "Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: Theoretical and empirical considerations" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY, WASHINGTON, DC, vol. 52, no. 12, 13 October 2006 (2006-10-13), pages 2194-2202, XP002470084 ISSN: 0009-9147 cited in the application	1-12, 14-18,31
A	page 2197, left-hand column, line 14 - page 2199, right-hand column, line 32 page 2201 figure 2	25,26,30
P,X	LO YMD ET AL: "Digital PCR for the molecular detection of fetal chromosomal aneuploidy." PNAS, vol. 104, no. 32, 7 August 2007 (2007-08-07), pages 13116-13121, XP007905909 page 13117, left-hand column - page 13118, right-hand column	1-12, 14-18, 25,26, 30,31
A	LO YMD ET AL: "Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection" NATURE MEDICINE, vol. 13, no. 2, February 2007 (2007-02), pages 218-223, XP007905910 cited in the application page 219, left-hand column, line 5 - line 25 page 219, right-hand column, line 26 - page 221, left-hand column page 222, right-hand column, line 13 - line 24	1-12, 14-18, 25,26, 30,31

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2008/002524

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EL KAROUI N ET AL: "Getting more from digital SNP data" STATISTICS IN MEDICINE, vol. 25, 2006, pages 3124-3133, XP007905913 cited in the application page 3125, line 25 - page 3128, line 19	1-12, 14-18, 25,26, 30,31
A	LO YMD AND CHIU RWK: "Prenatal diagnosis: progress through plasma nucleic acids." NATURE REVIEWS GENETICS, vol. 8, January 2007 (2007-01), pages 71-77, XP007905874 cited in the application the whole document	1-12, 14-18, 25,26, 30,31

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2008/002524

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.: 13,19-24,27-29
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:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
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:

see additional sheet

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

2. ☒ As all searchable claims could be searched without effort justifying an 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 reportcovers 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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 13,19-24,27-29

No opinion is given for claims 13, 19 to 24 and 27 to 29 because the subject-matter of said claims is not clearly defined contrary to the requirements of Article 6 PCT. The reasons are as follows:

According to claim 13 "the cutoff value is based on a measurement of one of the first portions or a measurement of one of the second portions". This vague expression does not enable the skilled reader to determine how the cutoff value is calculated nor which quantities are used in the calculation of said cutoff value. The subject-matter of claim 13 is unclear to such an extent that no meaningful search could be carried out.

According to claim 19, the "calculating of the first cutoff value includes at least one of: sequential probability ratio testing, false discovery rates, confidence intervals, and receiver operating characteristic curves." These listed methods allow for hypothesis testing (i.e. sequence imbalance vs sequence balance) based on given cutoff values. It is unclear how these methods are used for calculating the cutoff value itself.

Claims 20 to 24 define a proportion P1 and a set of probability values and state that the method includes "calculating the first cutoff value from the first proportion P1" (claim 20). This information, however, is insufficient to define said cutoff value. Moreover, the calculation of P1 itself is unclear: according to claim 21, P1 is calculated "using the first probability and the second probability". It is unclear how exactly these probabilities are used in the calculation of P1.

In claim 22, it is unclear what is being addressed by "an expected ratio relative to the non-reference nucleic acid sequence". Consequently, the calculation of the "first probability" remains unclear as well.

Claim 24 defines yet another probability, "third probability" but does not specify the purpose of this probability. It is unclear why this probability is defined and how it is used in the determination of the cutoff value. Also, according to claim 24 "calculating the second probability ... includes assuming that first probability and second probability are independent". It is unclear how the independence of the first and the second probabilities is incorporated in the calculation of the second probability.

The wording of claims 19 to 24 does not enable the skilled person to determine how the cutoff value is calculated. Consequently, no search regarding the definition of the cutoff value could be carried out. Although the description is more specific regarding the calculation of the cutoff values, the relevant passages are spread across pages 22 to 26 and could not be interpreted unambiguously.

Claim 27 also attempts to define the first cutoff value by "determining

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

a first average concentration ...; and multiplying the first average concentration by a factor derived from the first percentage to obtain a second average concentration ...". The claim only specifies that the factor is derived from the first percentage, but does not specify how exactly said factor is calculated. Furthermore, the result of the claim is a second average concentration. The claim does not specify how this second average concentration is used in the calculation of the cutoff value.

Claims 28 and 29 further specify the average concentration but do not enlighten the procedure of calculating the cutoff value.

Consequently, the method of calculating said cutoff value is unclear to such an extent that could not be meaningfully searched.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2)PCT declaration be overcome.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-12,14-18,31

A method for determining whether a nucleic acid sequence imbalance in a sample exists whereby the average concentration of a reference nucleic acid in a reaction is taken into account.

---

2. claims: 25-26,30

A method for determining whether a nucleic acid sequence imbalance in a sample exists whereby the sample consists of two or more types of cells and the percentage of the nucleic acid sequence from the first type of cells is taken into account.

---

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 March 2010 (11.03.2010)

PCT

(10) International Publication Number
WO 2010/027484 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12Q 1/25* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2009/004993

(22) International Filing Date:
4 September 2009 (04.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/094,843 5 September 2008 (05.09.2008) US
12/370,472 12 February 2009 (12.02.2009) US

(71) Applicant *(for all designated States except US)*: **PACIFIC BIOSCIENCES OF CALIFORNIA, INC.** [US/US]; 1505 Adams Drive, Menlo Park, CA 94025 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PATEL, Pranav** [IN/US]; 34203 Oneil Terrace, Fremont, CA 94555 (US). **KORLACH, Jonas** [DE/US]; 36100 Toulouse Street, Newark, CA 94560 (US). **BIBILLO, Arkadiusz** [PL/US]; 10100 Torre Ave, Apt. 144, Cupertino, CA 95014 (US). **BJORNSON, Keith** [US/US]; 8098 Peachtree Avenue, Newark, CA 94560 (US). **HANES, Jeremiah** [US/US]; 3243 Spring Street, Redwood City, CA 94063 (US).

(74) Agents: **LITTLEPAGE, Paul** et al.; Quine Intellectual Property Law Group, P.C., P.O. Box 458, Alameda, CA 94501 (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, 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, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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), European (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, 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: ENGINEERING POLYMERASES AND REACTION CONDITIONS FOR MODIFIED INCORPORATION PROPERTIES

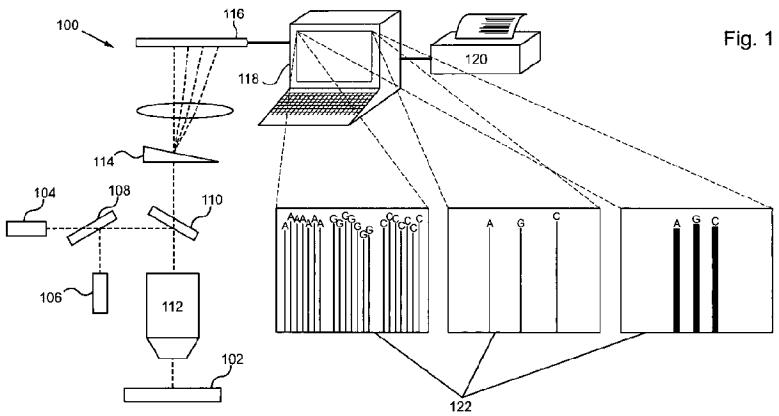


Fig. 1

(57) Abstract: Provided are methods for enhanced sequencing of nucleic acid templates. Also provided are reaction conditions that increase branching fractions during polymerization reactions. Also provided are compositions comprising modified recombinant polymerases that exhibit branching fractions that are higher than the branching fractions of the polymerases from which they were derived. Provided are compositions comprising modified recombinant polymerases that exhibit delayed translocation relative to the polymerases from which they were derived. Also provided are compositions comprising modified recombinant polymerases that exhibit increased nucleotide or nucleotide analog residence time at an active site of the polymerase. Provided are methods for generating polymerases with the aforementioned phenotypes and methods of using such polymerases to sequence a DNA template or make a DNA. Also provided are methods and nucleic acid sequencing systems for determining which labeled nucleotide is incorporated at a site during a template-dependent polymerization reaction.

WO 2010/027484 A2

ENGINEERING POLYMERASES AND REACTION CONDITIONS FOR MODIFIED INCORPORATION PROPERTIES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to Provisional U.S. Patent Application No. 61/094,843, filed September 5, 2008, and U.S. Patent Application No. 12/370,472, filed February 12, 2009, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

FIELD OF THE INVENTION

[0002] The invention is in the field of nucleic acid sequencing, for example, single molecule sequencing.

BACKGROUND OF THE INVENTION

[0003] High throughput sequencing has become a central tool in the field of biotechnology and is revolutionizing personalized medicine. Many diseases and/or disorders are genetic in origin. Acquiring the genomic sequence of individual patients in a comprehensive, rapid and cost-effective manner enhances the ability of medical professionals to diagnose diseases or identify predispositions to diseases or other genetic-based disorders. Genomic sequence information also enhances the treatment of diseases by providing doctors with information regarding the efficacy of a given therapy for a particular individual.

[0004] One approach aimed at efficiently obtaining the complete genomic sequence of an organism is sequencing by incorporation, where the identity of the sequence of nucleotides in a template nucleic acid polymer is determined by identifying each complementary base that is added to a nascent strand being synthesized against the template sequence, as such bases are added. While detection of added bases may be a result of detecting a byproduct of the synthesis or extension reaction, e.g., detecting released pyrophosphate, in many systems and processes, added bases are labeled with fluorescent dyes that permit their detection. By uniquely labeling each base with a distinguishable fluorescent dye, one attaches a distinctive

detectable characteristic to each dye that is incorporated, and as a result provides a basis for identification of an incorporated base, and by extension, its complementary base upon the template sequence.

[0005] During sequencing by incorporation, nucleotide (or nucleotide analog) incorporation events are detected in real-time as the bases are incorporated into the extension product. This can be accomplished by immobilizing the complex within an optically confined space or otherwise resolved as an individual molecular complex. Some sequencing by incorporation methods employ nucleotide analogs that include fluorescent labels coupled to the polyphosphate chain of the analog, which are then exposed to the complex. Upon incorporation, the nucleotide – along with its fluorescent label – is retained by the complex for a time and in a manner that permits the detection of a signal "pulse" from the fluorescent label at the incorporation site. Upon completion of incorporation, all but the alpha phosphate group of the nucleotide is cleaved away, liberating the label from retention by the complex, and diffusing the signal from that label.

[0006] Thus, during an incorporation event, a complementary nucleotide analog, including its fluorescent label, is effectively "immobilized" for a time at the incorporation site, and the fluorescent label is subsequently released and diffuses away when incorporation is completed. Detecting the localized "pulses" of fluorescent tags immobilized at the incorporation site, and distinguishing those pulses from a variety of other signals and background noise, allows bases to be called in real-time as they are incorporated. Further details regarding base calling during sequencing by incorporation methods are found in Tomaney et al. PCT Application Serial No. PCT/US2008/065996 METHODS AND PROCESSES FOR CALLING BASES IN SEQUENCING BY INCORPORATION METHODS, incorporated herein by reference in its entirety for all purposes.

[0007] Current real-time sequencing by incorporation methods may exhibit sub-optimal reliability and accuracy due to missed signal pulses that contribute as errors in sequencing reads. Missed pulses derive from, e.g., insufficient residence time of the analogs at an active site of the polymerase or unlabeled or broken-fluorophore nucleotide analogs. Compositions

and methods for improving the reliability and accuracy of sequencing by incorporation are desirable.

SUMMARY OF THE INVENTION

[0008] Altered reaction conditions and modified DNA polymerases can find use in such applications as, e.g., single-molecule sequencing (SMS), genotyping analyses such as SNP genotyping using single-base extension methods, and real-time monitoring of amplification, e.g., real time PCR. The invention provides methods of sequencing a nucleic acid template, which methods utilize signal pulses or signatures from branch fraction nonincorporation events (and, optionally, actual nucleotide incorporation events) to determine which nucleotide is incorporated at a particular site/position of the template nucleic acid. The invention further provides methods that modulate (e.g., increase) the branching rate of a polymerization reaction to facilitate identifying which nucleotide is incorporated at a particular site. A nucleic acid sequencing system that detects and utilizes signal pulses or signatures from branch fraction nonincorporation events to determine the sequence of a template nucleic acid is also provided by the invention. The invention further provides compositions that include modified recombinant polymerases that exhibit properties, e.g., increased branching fraction, delayed translocation or increased nucleotide or nucleotide analog residence time, which can be particularly desirable for these applications. These improved polymerase properties can facilitate readout accuracy. In addition, the invention provides methods of generating the modified polymerases of the invention and methods in which such polymerases can be used to e.g., sequence a DNA template and/or make a DNA.

[0009] In one aspect, the invention provides methods for determining which labeled nucleotide is incorporated at a particular site during a template dependent polymerization reaction. The methods include incorporating the nucleotide into a nucleic acid polymer, whereby signal pulses or signatures are generated from branch fraction nonincorporation events and, optionally, actual nucleotide incorporation events for the site. The methods additionally include monitoring a time course of at least branch fraction signal pulses or signatures produced by the polymerization reaction and assigning which nucleotide is incorporated at the site, using at least signal pulses or signatures from branch fraction nonincorporation sampling

events. The methods optionally comprise counting or estimating the number of redundant iterative sampling signal pulses per incorporation event, or determining an average number of redundant signal pulses per incorporation event. Optionally, the polymerization reaction is a high branch fraction polymerization reaction, where the branch fraction is optionally 70% or more, 80% or more, or 90% or more.

[0010] The methods described above optionally include at least one species of metal ion, which metal ion increases the frequency of branch fraction nonincorporation events in the reaction. Example metal ions include: Mg^{++}, Mn^{++}, Zn^{++}, Co^{++}, Ca^{++}, Fe^{++}, Cr^{++} and Sr^{++}. The methods described above optionally comprise both Mg^{++} and Mn^{++}, e.g., where the concentration of Mg^{++} is higher than the concentration of Mn^{++}.

[0011] The methods described above optionally include a Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, T4 or T7 DNA polymerase, or a modified recombinant DNA polymerase thereof. The modified recombinant polymerase can optionally exhibit a higher branching fraction as compared to a corresponding wild-type polymerase, or an increased exonuclease rate that is about 10% to 50% as compared to its polymerization rate.

[0012] The methods described above optionally include branch fraction nonincorporation events that comprise iterative sampling of labeled unincorporatable nucleotide analogs, optionally including actual nucleotide incorporation events that comprise incorporation of unlabeled nucleotides. Actual nucleotide incorporation events optionally include incorporation of nucleotides that are differentially labeled as compared to the unincorporatable nucleotides. In one example, labeled unincorporatable nucleotide analogs comprise a link between an alpha and beta phosphate group that is not hydrolyzable by a polymerase enzyme.

[0013] The methods described above optionally include a polymerase enzyme, polymerase reaction conditions, and/or polymerase reaction substrates that are selected such that the polymerization reaction exhibits two kinetically observable steps within an observable phase of the polymerase reaction. The two kinetically observable steps are optionally steps which proceed in a bright phase or a dark phase, and the polymerase enzyme optionally

comprises a modified recombinant Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, T4 or T7 DNA polymerase. Optionally, the polymerase reaction conditions can include, e.g., a selected metal cofactor concentration, a selected pH, a selected temperature, an enzyme activity modulator, D_2O, an organic solvent, and a buffer.

[0014] The methods described above optionally comprise branch fraction nonincorporation events that comprise noncognate branch fraction nonincorporation events. Branch fraction signal pulses or signatures are optionally generated from noncognate branching events of a nucleotide analog, e.g., guanine and thymine.

[0015] The reaction of the methods described above is optionally reacted in a DNA sequencing system, where the DNA sequencing system optionally comprises a zero mode waveguide or nanohole.

[0016] Assigning the nucleotide in the methods described above optionally comprises applying a statistical model to the signal pulses or signatures generated from branch fraction nonincorporation events, signal pulses generated from actual nucleotide incorporation events, or both, which statistical model assigns a likelihood that the signal pulses or signatures correspond to an incorporation event.

[0017] The methods described above optionally comprise performing an additional template dependent polymerization reaction under high processivity reaction conditions, monitoring a second time course of signal pulses or signatures produced by the additional polymerization reaction, and compiling sequencing information derived from the second time course of signal pulses or signatures with sequencing information derived from the time course of branch fraction signal pulses or signatures.

[0018] In another aspect, the invention provides multi-modal sequencing methods that comprise performing a first template dependent sequencing reaction in a first mode comprising a first set of reaction conditions and collecting initial sequencing information produced by the first sequencing reaction. Additionally, the methods can include performing a second sequencing reaction of the template, or a copy thereof, in a second mode that includes a second

set of reaction conditions and collecting additional sequencing information produced by the second sequencing reaction. The methods can include compiling the initial and additional sequencing information to provide a sequence of at least a portion of the template.

[0019] The second sequencing reaction of the methods described above is optionally produced by altering one or more reaction conditions of the first sequencing reaction, and the initial and additional sequencing information are collected in real time. Optionally, altering one or more reaction conditions comprises adding one or more polymerase cofactors to the first sequencing reaction, where the cofactors of the first sequencing reaction are optionally Mn^{++} or Mg^{++}. Both the first and second sequencing reactions comprise single template molecule sequencing reactions. For example, the first mode can produce a higher branch fraction than the second mode. The second mode optionally produces longer read lengths than the first mode. The template of the methods described above is optionally adapted to sequencing, e.g., a single circular template molecule, e.g., where the method includes switching between the first and second modes.

[0020] The invention also provides methods for determining which of two or more labeled nucleotides is incorporated at a site during a template-dependent polymerization reaction. The methods include incorporating the nucleotide into a nucleic acid polymer produced by the polymerization reaction, whereby signal pulses or signatures are generated. The methods further include monitoring the pulses or signatures, and using the presence of multiple pulses corresponding to the nucleotide, or identical molecules thereof, to assign which labeled nucleotide is incorporated at the site. The multiple pulses optionally include 2 to 20 pulses and can be generated from branch fraction nonincorporation events, which events are optionally induced by sequencing compositions that include at least one species of metal ion. Metal ions of the method can include Mg^{++}, Mn^{++}, Zn^{++}, Co^{++}, Ca^{++}, Fe^{++}, Cr^{++} and Sr^{++}. Optionally, using the presence of multiple pulses comprises distinguishing incorporation and nonincorporation signals to assign which labeled nucleotide was incorporated at the site.

[0021] In another aspect, the invention provides nucleic acid sequencing systems that, during operation of the system, sequences a nucleic acid. The nucleic acid sequencing system comprises a signal detector that detects at least signal pulses or signatures from branch fraction nonincorporation events during sequencing of a template nucleic acid, system instructions or

software that assigns a sequence based upon detection of at least signal pulses or signatures from branch fraction nonincorporation events, and a user output module that displays the sequence to the user.

[0022] Signal pulses or signatures from branch fraction nonincorporation events of the nucleic acid sequencing systems described above are optionally generated during a first sequencing mode, where the signal detector subsequently detects signal pulses generated during a low branch fraction second sequencing mode, and where the system instructions assign a sequence based upon detection of signal pulses or signatures from the first and second sequencing modes.

[0023] The nucleic acid sequencing systems optionally comprise a zero-mode waveguide or nanohole proximal to the signal detector, where during operation of the system, a sequencing reaction is contained by the zero-mode waveguide or nanohole.

[0024] In another aspect, the invention provides compositions that include a modified recombinant nucleic acid polymerase that exhibits an altered property selected from an increased branching fraction during a polymerization by the polymerase, an altered translocation property of the polymerase during a polymerization reaction, and a combination of these two altered properties, where the altered property or properties is altered as compared to a corresponding wild-type polymerase.

[0025] The modified recombinant polymerase of the compositions described above can optionally be a modified recombinant Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, T4, or T7 polymerase. Other available polymerases can also be used as starting points for modification to alter translocation rates or to modulate branch fraction activity, such as reverse transcriptases and DNA-dependent RNA polymerases.

[0026] The modified recombinant polymerase exhibiting an increased branching fraction can optionally comprise at least one amino acid substitution or deletion or combination of substitutions or deletions selected from: N62D and Y454A; D362S; Y259H; F237Y; L381I; Y369H; H461Y; A377G; K138Q; H461D; A377S; N62D and K371Q; V118L; and K124R; where numbering of the residues is relative to a wild-type Φ29 polymerase of SEQ ID NO: 3.

The modified recombinant polymerase exhibiting an increased branching fraction can optionally exhibit increased exonuclease activity, where the polymerase exhibits an exonuclease rate that is about 10% to 50% as compared to its polymerization rate.

[0027] Modified recombinant polymerases of the compositions described above optionally exhibit a branching fraction that is at least 50% greater, at least 100% greater, or at least 200% greater than the branching fraction of a wild-type Φ29 polymerase of SEQ ID NO: 3. Optionally, the polymerases exhibit increased exonuclease activity as compared to the corresponding wild type polymerase, where the increased exonuclease activity is optionally about 10% to 50% as compared to its polymerization rate.

[0028] The modified recombinant polymerase exhibiting an altered translocation property can optionally comprise a fusion protein that comprises at least a subsequence of the parental polymerase (e.g., a Φ29 DNA polymerase) and at least one heterologous polypeptide sequence (see, e.g., SEQ ID No. 1 and SEQ ID No. 2 in the sequence listing herein). Optionally, the fusion of the at least a subsequence of the parental polymerase and the heterologous polypeptide sequence can occur at or near the c-terminal end of the parental polymerase. The wild-type polymerase is optionally a Φ29 polymerase. The heterologous polypeptide sequence can optionally comprise at least one charged amino acid, where the at least one charged amino acid can optionally be histidine or a chain of histidines. Optionally, the fusion proteins described above can comprise a linker between the at least a subsequence of the parental polymerase and the heterologous polypeptide sequence, where the linker optionally comprises a Ser3Gly linker.

[0029] The modified recombinant polymerase exhibiting an altered translocation property can optionally comprise at least one amino acid substitution or deletion or combination of substitutions or deletions selected from Asp570Lys; Asp570Ala; Asn313Lys; Asn313Ala; Gln303Lys; Gln303Ala; Gly532Ser; Met533delet; Cys530delet; Met533delet and Cys530delet; Gly532delet; Ala531Gly; Thr573Lys; Thr573Ala; Asn396Lys; Thr571Lys; Thr571Ala; Thr534Lys; Thr534Ala; Asp535Lys; Asp535Ala; Lys529Ala; and Lys529Asn; where numbering of the residue positions is relative to a wild-type Φ29 polymerase of SEQ ID NO: 3.

[0030] The altered translocation property of the modified recombinant DNA polymerases can optionally comprise a delay in translocation. Modified recombinant polymerases of the compositions described above can optionally exhibit a delay in translocation that is at least about 2.5x, 10x or 15x greater than a corresponding wild-type polymerase.

[0031] The modified recombinant polymerases of the compositions described above optionally exhibit an increased nucleotide or nucleotide analog residence time or increased processivity as compared to a corresponding wild-type polymerase.

[0032] The compositions comprising a modified recombinant polymerase that exhibits an altered property described above can include a phosphate-labeled nucleotide analog, a DNA template, and a modified recombinant DNA polymerase, e.g., any of the polymerases described above, that can incorporate the nucleotide analog into a copy nucleic acid in response to the DNA template. These compositions can be present in a DNA sequencing system, e.g., a zero-mode waveguide or nanohole. Optionally, the polymerase of the compositions can be immobilized on a surface.

[0033] In a related aspect, the invention provides methods of sequencing a nucleic acid template. The methods include providing a reaction mixture that includes the nucleic acid template, a replication initiating moiety that complexes with or is integral to the template, the modified recombinant nucleic acid polymerase of the compositions described above, where the polymerase is capable of replicating at least a portion of the template using the moiety in a template-dependent polymerization reaction, and one or more nucleotides and/or nucleotide analogs. In addition, the methods subject the reaction mixture to a polymerization reaction in which the modified recombinant polymerase replicates at least a portion of the template in a template-dependent manner, where one or more nucleotides and/or nucleotide analogs are incorporated into the resulting copy nucleic acid. The methods additionally identify a time sequence of incorporation of the one or more nucleotides and/or nucleotide analogs into the resulting copy nucleic acid. Optionally, the methods include a modified recombinant polymerase that exhibits increased processivity relative to the wild-type polymerase. The methods optionally include identifying the time sequence of incorporation by observing more than one signal pulse per nucleotide incorporation event. Subjecting the reaction mixture to a

polymerization reaction and identifying a time of sequence incorporation can optionally be performed in a zero mode waveguide, nanohole or other micro- or nano-structure.

[0034] The invention also provides methods of making a nucleic acid that include providing a reaction mixture that comprises a template, a replication initiating moiety that complexes with or is integral to the template, a modified recombinant DNA polymerase with an altered property or combination of altered properties, e.g., such as those described above, which can replicate at least a portion of the template using the moiety in a template-dependent polymerase reaction, and one or more nucleotides and/or nucleotide analogs. In addition, the methods include reacting the mixture such that the polymerase replicates at least a portion of the template in a template-dependent manner, whereby the one or more nucleotides and/or nucleotide analogs are incorporated into the resulting nucleic acid. Optionally, the methods include detecting incorporation of at least one of the nucleotides and/or nucleotide analogs, which optionally includes observing more than one signal pulse per nucleotide incorporation event. The mixture is optionally reacted in a zero mode waveguide or nanohole, and the modified recombinant polymerase optionally exhibits an increased nucleotide or nucleotide analog residence time and/or processivity as compared to the parental polymerase.

[0035] In a related aspect, the invention provides methods of making a modified recombinant DNA polymerase that include mutating a polymerase of interest, e.g., a Φ29-type DNA polymerase, and selecting resulting modified polymerases for a property selected from increased branching fraction and altered translocation. Mutating the polymerase of interest can optionally comprise structurally modeling the polymerase to identify a feature that may affect branch fraction or altered translocation. Optionally, mutating the polymerase of interest includes making a library of modified recombinant polymerases, and selecting the modified polymerases includes screening the library to identify at least one member exhibiting the property. The polymerase of interest optionally includes a Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, T4 or T7 polymerase. Modified recombinant polymerases that exhibit the property optionally exhibit increased nucleotide or nucleotide analog residence time, or increased processivity, as compared to a corresponding wild-type polymerase.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 schematically illustrates a system of the invention.

Figure 2 schematically illustrates a sequencing by incorporation reaction and the resulting characteristics of signal pulses detected by a system that employs reaction conditions or polymerases of the invention.

Figure 3 is a time sequence of signal pulses generated from a sequencing by incorporation reaction under relatively high branch fraction conditions.

Figure 4 is a schematic illustration of a sequencing by incorporation reaction in which unincorporatable nucleotides are included in the reaction.

Figure 5 shows a theoretical representation of the frequency of binding events per incorporation for a polymerase reaction having one rate-limiting step or two rate-limiting steps within an observable phase.

Figure 6 is a schematic illustration of the reaction cycle for polymerase-mediated nucleic acid primer extension.

Figure 7 schematically illustrates a system of the invention that utilizes more than one mode of sequencing.

DETAILED DESCRIPTION

[0036] The invention is generally directed to modified or engineered compositions that are characterized by modified profiles or characteristics for incorporation of nucleotides in template directed nucleic acid synthesis. Such characteristics include, for example, increased frequency of branching events, changes in reaction rates that lead, e.g., to delayed polymerase translocation and/or increased nucleotide or nucleotide analog retention time during polymerization reactions. Individually or in combination, these modifications can increase sequence readout accuracy (e.g., increase sequence accuracy in single molecule sequencing reactions) using the methods of the invention. Polymerases of the invention optionally also include additional mutations or modifications that provide other desirable features, e.g., modify one or more kinetic features of the polymerase (e.g., increased processivity), increased surface stability for polymerases bound to a surface, or the like.

[0037] During sequencing by incorporation, e.g., single molecule sequencing by synthesis (SMS), nucleotide (or nucleotide analog) incorporation events are detected in real-time as the bases are incorporated into the extension product. This can be accomplished by immobilizing a synthesis complex, which includes a polymerase enzyme, such as a DNA polymerase enzyme, a template nucleic acid sequence, and a primer sequence that is complementary to a portion of the template sequence, within an optically confined space or otherwise resolved as an individual molecular complex. Some SMS methods employ nucleotide analogs that include fluorescent labels coupled to the polyphosphate chain of the analog, which are then exposed to the complex. Upon incorporation, the nucleotide – along with its fluorescent label – is retained by the complex for a time and in a manner that permits the detection by a sequencing system of a signal "pulse" from the fluorescent label at the incorporation site. The sequentially detected signal pulses are then interpreted by the sequencing system to generate a readout corresponding to the sequence of the template nucleic acid. For a discussion of preferred sequence by incorporation processes, see, e.g., U.S. Patent Nos. 6,056,661, 7,052,847, 7,033,764, 7,056,676, 7,361,466, the full disclosures of which are hereby incorporated herein by reference in their entirety for all purposes. Further details regarding base calling during sequencing by incorporation methods are found in Tomaney et al. PCT Application Serial No. PCT/US2008/065996 METHODS AND PROCESSES FOR CALLING BASES IN SEQUENCING BY INCORPORATION METHODS, incorporated herein by reference in its entirety for all purposes.

[0038] **Figure 2** schematically illustrates a sequencing by incorporation reaction and the resulting patterns of signal pulses detected by a system that employs reaction conditions or polymerases of the invention. Figure 2A schematically illustrates a polymerization reaction where dye-labeled nucleotides are incorporated in a stepwise fashion according to the sequence of the template strand. When a dye-labeled nucleotide enters the detection region (dashed box) which encompasses the polymerase, the dye emits optical signal pulses or signatures in response to excitation radiation that are detected by a signal detector. Examples of detection methods and optically confined reaction regions include, e.g., Zero Mode Waveguides, e.g., as described in U.S. Patent Nos. 6,917,726, 7,013,054, 7,181,122, and 7,292,742, the full disclosures of which are hereby incorporated by reference in their entirety for all purposes.

Figure 2B schematically illustrates the patterns or characteristics of signal pulses that would arise under standard conditions (Panel I), increased branching conditions (Panel II), conditions that include a polymerase with a decreased translocation rate (Panel III) and conditions that include a polymerase that exhibits increased nucleotide analog residence time (Panel IV). The resulting patterns or characteristics of signal pulses from the various conditions are described in detail below.

I. INCREASED BRANCHING

[0039] "Branching" is a phenomenon that occurs during polymerization. During a polymerase kinetic cycle, sampling of each of four possible nucleotides (or nucleotide analogs) occurs until a correct Watson-Crick pairing is generated (see, e.g., Hanzel et al. WO 2007/076057 POLYMERASES FOR NUCLEOTIDE ANALOG INCORPORATION for an example model description of the kinetic cycle of a polymerase). However, chemical linkages between a sampled nucleotide and a 3'OH group of a preceding base can fail to occur for a correctly paired nucleotide, due, e.g., to release of the correctly paired base from the active site. This can occur as a result of the nucleotide leaving the site without a covalent bond being formed, or e.g., as a result of cleavage of the covalent bond (e.g., due to exonuclease activity) prior to polymerase translocation to the next incorporation site. During single molecule sequencing (SMS) procedures, and particularly those single molecule processes that monitor incorporation in real time, where both the failed incorporation and the actual incorporation of the nucleotides provide signal pulses, sequences deciphered for the incorporation site can have an incorrect "insertion" relative to the correct sequence as a result of such branching. This phenomenon is termed "branching" because it leads to a branch in the sequence (a site where two identical molecules will be read as having different sequences) and may lead to increased error rates during SMS.

[0040] While branching can, in many applications of single molecule sequencing processes, be viewed as an accuracy reducing phenomenon, in at least a first aspect of the present invention, increased branching is exploited to increase sequence accuracy by providing redundant signal events resulting from iterative sampling of labeled nucleotides or

nucleotide analogs. In particular, improved sequence reliability and accuracy is achieved by providing reaction conditions and/or polymerases that exhibit a relatively high average branching fraction for a particular nucleotide or nucleotide analog and a certain distribution of branch signal pulses around this average. Such compositions are used in combination with a sequencing system that observes and interprets more than one signal pulse or signature per incorporation event to identify the nucleotide sequence of a target or template nucleic acid. This is advantageous in the present invention because detecting more than one signal pulse or signature per incorporation event provides inherent redundancy of signal for each desired incorporation event. In some cases, the "signature" will include regions of optical signal versus time that are characteristic of the branching nucleotide, but do not appear as individual pulses. This can occur, for example, when a sequence of pulses are not individually resolved. Further details regarding sequencing under high branch fraction conditions can be found in Bjornson et al. PCT Application Serial No. PCT/US2009/000921 COMPOSITIONS AND METHODS FOR USE IN ANALYTICAL REACTIONS, incorporated herein by reference in its entirety for all purposes. Additional information useful to sequencing under high branch fraction conditions can be found in Bjornson et al. PCT Application Serial Number PCT/US2009/002003 TWO SLOW-STEP POLYMERASE ENZYME SYSTEMS AND METHODS, incorporated herein by reference in its entirety for all purposes.

[0041] The branching fraction is the proportion of cognate nucleotide (or nucleotide analog, e.g., A488dA4P) dissociation events from the polymerase active site as compared to the total number of events, e.g., the sum of the dissociation events and the incorporation events for the cognate nucleotide or nucleotide analog. The present invention provides high branch fraction polymerization reactions. As used herein, a high branch fraction polymerization reaction includes a reaction that exhibits a branching fraction of at least about 70% or more, about 80% or more, about 85% or more, about 90% or more, or about 95% or more. For example, in a sequencing reaction in which the branching fraction is 80%, 80% of the total interactions of the nucleotide or nucleotide analog with the polymerase binding pocket result in dissociation, rather than incorporation, of the nucleotide or nucleotide analog.

[0042] An aspect of the invention is a method of nucleic acid sequencing by monitoring an optical signal from a polymerase reaction, wherein the base call, or the assignment of the incorporated base is made on the basis of multiple pulses from the same nucleotide. The number of pulses used to assign which base has been incorporated may depend on the branching fraction under the conditions of the polymerase reaction. In some cases, the number of pulses used to assign which nucleotide has been incorporated will vary between the different nucleotides in that reaction medium. The number of pulses used to assign which nucleotide is incorporated can be about 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, or more pulses. The number of pulses can be between about 2 and about 30 pulses, between about 2 and about 20 pulses, or between about 3 and about 25 pulses. The number of pulses can be expressed, for example, as the average number of pulses used to assign a given nucleotide for one sequencing reaction.

[0043] Sequence read errors during SMS can also derive from the incorporation of nucleotides or nucleotide analogs that constitute dark matter (for the purposes of this disclosure, "dark matter" refers to unlabeled nucleotides or nucleotide analogs with nonfunctional labels, e.g., broken fluorophores). Here, a genuine incorporation event is not detected due to the absence of a signal pulse from the dark matter, and a subsequent incorporation event is interpreted by the sequencing system as occurring at the position where the dark matter was incorporated. Dark matter, therefore, may potentially contribute to error rates in single molecule sequencing that utilizes the incorporation of labeled nucleotides.

[0044] In certain aspects, the reaction conditions, modified recombinant polymerases, and/or nucleotide analogs of the present invention – employed in conjunction with the sequencing system of the present invention – reduce sequence read errors that might result from missed pulses. The reaction conditions induce a relatively high branch fraction polymerization reaction – and the modified recombinant polymerases exhibit increased average branching fractions - such that a greater number of nucleotide analogs, which, if incorporated, would correctly pair with the corresponding nucleotide of the template strand, enter the active site before an analog is eventually incorporated into the extension product. The nucleotide analogs that enter the active site, but fail to incorporate, produce redundant signal pulses or signatures at each incorporation site, resulting in multiple redundant signal

events for each incorporation event. An example of signal pulses generated under relatively high branch fraction conditions is shown in **Figure 3.**

[0045] The sequencing system takes into account the average branching fraction of the polymerase and a certain distribution of branch pulses or signatures per nucleotide incorporation around this average. Because multiple signal pulses are observed for each incorporation event, branching events involving unlabeled nucleotides or nucleotides with nonfunctional labels, i.e., dark matter, do not result in a sequencing read error, but rather only slightly decrease the distribution of the average number of pulses or signatures per incorporation. In the event that dark matter is incorporated into the extension product, signal pulses derived from branching events involving nucleotide analogs with functional labels prior to dark matter incorporation can provide sufficient redundancy for determining the correct base at the incorporation site.

[0046] As will be appreciated, high branch fraction sequencing conditions can also be used for sequencing RNA templates, for example using reverse transcriptase enzymes and for RNA synthesis, for example by DNA dependent RNA polymerases.

A. Enhanced Sequencing Using Reaction Conditions That Promote Branching

[0047] The present invention provides reaction conditions – such as the type, level, and relative amounts of cofactors – that increase the frequency of branching events during nucleic acid polymerization reactions. Such reaction conditions may be used in combination with polymerases that are engineered to exhibit increased branching fractions under selected conditions, or can be used with polymerases that are unaltered with respect to branching properties. The phosphoryl transfer reaction of DNA polymerases is typically catalyzed by a two-metal ion mechanism, where two divalent metal ions, e.g., Mg^{++} and/or Mn^{++}, complexed with the DNA polymerase facilitate the incorporation of a nucleotide into the 3'OH of the extension product. One of the metal ions is proposed to interact with the 3'OH of the primer strand, thereby facilitating its attack on the α-phosphate of the incoming nucleotide. Both metal ions are believed to stabilize the transition state that occurs during the course of the extension reaction.

[0048] During the course of the polymerase reaction, divalent metal cofactors, such as magnesium or manganese, will interact with the enzyme-substrate complex, playing a structural role in the definition of the active site. For a discussion of metal cofactor interaction in polymerase reactions, see, e.g., Arndt, et al., Biochemistry (2001) 40:5368-5375. For example, and without being bound to any particular theory of operation, it is understood that metal cofactor binding in and around the active site serves to stabilize binding of incoming nucleotides. For further details regarding the effect of metal cofactors on polymerase kinetics and nucleic acid synthesis reactions, see Bjornson et al. PCT Application Serial Number PCT/US2009/002003 TWO SLOW-STEP POLYMERASE ENZYME SYSTEMS AND METHODS, incorporated herein by reference in its entirety for all purposes.

[0049] In the context of the present invention, it has been discovered that modulation of the concentration of a divalent metal cofactor, or competitive modulation of two or more divalent metal cofactors, to the synthesis reaction can result in increased branching for enhanced nucleic acid sequencing without a consequent increase in negative reaction events. As described in detail herein, the increased branching provides redundant signal pulses or signatures, thereby reducing or eliminating the occurrence of missed signal pulses and improving sequence accuracy. As used herein, a signature can include regions of optical signal versus time that is characteristic of the branching nucleotide, but does not appear as an individual signal pulse.

[0050] In the synthesis reaction, certain divalent or trivalent metal cofactors, such as magnesium and manganese are known to interact with the polymerase to modulate the progress of the reaction (See, e.g., U.S. Patent No. 5,409,811). As will be appreciated, depending upon the nature of the polymerization reaction, environmental conditions (e.g., temperature, pH, etc.), the polymerase used, the nucleotides employed, etc., different metal co-factors will have widely varying catalytic effects upon the polymerization reaction. In the context of the present invention, different metal cofactors will be referred to herein based upon their relative catalytic impact on the polymerization reaction, as compared to a different metal included under the same reaction conditions. For purposes of discussion, a first metal cofactor that interacts with the polymerase complex to support the polymerization reaction to a higher level than a second metal cofactor under the same conditions is termed a "catalytic metal ion" or "catalytic metal."

[0051] The present invention provides sequencing compositions and methods that include, e.g., divalent metal ions at concentrations that induce high branch fraction polymerization reactions. Divalent metal ions of the invention can be, e.g., Mg^{++}, Mn^{++}, Zn^{++}, Co^{++}, Ca^{++}, Fe^{++}, Cr^{++}, and/or Sr^{++}. For the purposes of this disclosure, a high branch fraction polymerization reaction includes a reaction that exhibits a branching fraction of about 70% or more, about 80% or more, about 85% or more, about 90% or more, or about 95% or more.

[0052] In one embodiment, a high branch fraction polymerization reaction is induced in a sequencing composition that includes a DNA polymerase, e.g., a Φ29-derived polymerase that uses Mn^{++} as the sole source of metal cofactor at a concentration of about 250 μM or less, about 200 μM or less, about 150 μM or less, about 125 μM or less, about 100 μM or less, about 75 μM or less, or about 50 μM or less. By contrast, branching is typically not promoted in sequencing compositions that include, e.g., about 500 μM Mn^{++} or more, in the absence of other factors.

[0053] In another embodiment, a high branch fraction polymerization reaction is induced by a sequencing composition that includes a DNA polymerase, e.g., a Φ29-derived polymerase that uses Mg^{++} as the sole of metal cofactor at a concentration of about 1 mM or more, about 2 mM or more, about 3 mM or more, about 5 mM or more, about 10 mM or more, about 20 mM or more, about 30 mM or more, about 40 mM or more, or about 50 mM or more.

[0054] In yet another embodiment, a high branch fraction polymerization reaction is induced by a sequencing composition that includes a DNA polymerase, e.g., a Φ29-derived polymerase that uses both Mg^{++} and Mn^{++} as metal cofactors, where both the absolute values and ratio of the two concentrations of Mg^{++} and Mn^{++} determines the extent of branching. For example, the present invention provides sequencing compositions that induce branching by including both Mg^{++} and Mn^{++}, where the concentration of Mg^{++} is greater than the concentration of Mn^{++}. In one particular embodiment, the sequencing composition includes about 10 mM $MgCl_2$ and about 100 μM $MnCl_2$. A range of suitable concentrations to increase branching are 0-200 mM $MgCl_2$ and 0.01-50 mM $MnCl_2$, and all possible combinations of values between those two ranges.

[0055] In another embodiment, the reaction conditions include Mn^{++} and a metal cofactor other than Mg^{++}, such as calcium, barium, strontium, iron, cobalt, nickel, tin, zinc, and europium. For example, these metals can be added to the polymerization reaction in salt form such as $Sr(OAc)_2$, $CoCl_2$, $SnCl_2$, $CaCl_2$, or $ZnSO_4$. Both the absolute values and ratio of the two concentrations can influence the extent of branching.

[0056] The present invention also provides methods for increasing the branching fraction during sequencing by incorporation by modifying reaction conditions other than the identity and/or concentrations of metal ions. For example, the pH (lowering the pH to about 6.5), temperature (e.g., decreased temperature), addition of D_2O, and/or addition of small molecule inhibitors (e.g., a noncompetitive inhibitor that slows covalent attachment of the nucleotide to the 3-OH of the growing strand of the nucleic acid being synthesized, e.g., a non-competitive HIV-RT inhibitor), can be used to alter the branching fraction of the polymerization reaction.

[0057] As will be appreciated, the sequencing compositions and methods described above that utilize metal cofactors to induce relatively high branching fractions can be used in combination with any other embodiments described herein, including: (1) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (2) iterative sampling of unincorporatable nucleotides; (3) two slow-step enzyme systems; (4) detection of noncognate branching events; (5) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (6) modified recombinant polymerases that exhibit increased nucleotide residence time; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein. As will also be appreciated, high branch fraction sequencing conditions can be used for sequencing RNA templates, for example using reverse transcriptase enzymes, and for RNA synthesis, for example by DNA dependent RNA polymerases.

B. Enhanced Sequencing Using Modified Recombinant Polymerases with Increased Branching Fractions

[0058] During a polymerase kinetic cycle, sampling of each of the possible nucleotides or nucleotide analogs occurs until a correct Watson-Crick pairing is generated. According to structural studies of DNA polymerases complexed with DNA substrates, the primer-terminus does not typically form a covalent bond with an incorrectly paired nucleotide (Berman, et al. (2007) "Structures of phi29 polymerase complexed with substrate: the mechanism of translocation in polymerases." *EMBO J* **26:** 3494–3505). Conversely, branching events can occur during the polymerase kinetic cycle, where chemical linkages between a correctly paired nucleotide and a 3'OH of a preceding base fail to form, e.g., due to premature release of the sampled nucleotide from the active site. The kinetic cycle is then repeated for the same site, eventually resulting in the physical incorporation of the correct nucleotide.

[0059] As described above, these branching events can result in sequence read errors in standard sequencing methods, e.g., due to extra incorporation signals relative to the template sequence, received by a sequencing system that monitors signal pulses from the nucleotide analog at the active site as a proxy for incorporation, if the system does not account for the branching events. However, a sequencing system that utilizes branching events and calls bases according to multiple signal pulses or signatures for each incorporation event can be used in combination with polymerases that exhibit a high average branch fraction to improve sequence read accuracy. Under such conditions, redundant signals pulses or signatures resulting from iterative sampling of a labeled cognate nucleotide or nucleotide analog can reduce the error rate of a sequence read, as compared to lower branch fraction sequencing conditions where only one or a small number of signal pulses for each incorporation event can go undetected by the sequencing system.

[0060] The present invention provides modified recombinant polymerases and reaction conditions with increased branching fractions that can be used to improve sequence read accuracy. The branching fraction is the proportion of cognate nucleotide (or nucleotide analog, e.g., A488dA4P) dissociation events from the polymerase active site as compared to the total number of events, e.g., the sum of the incorporation events and dissociation events for the cognate nucleotide or nucleotide analog. Either incorporation or non-incorporation events, or both, can be detected by monitoring a signal profile produced by a sequencing reaction.

[0061] In the present invention, modification of a DNA polymerase by mutagenesis is used to increase the frequency of branching events. In exemplary embodiments, this modification may include one or more of either creating a more loosely structured binding pocket for the (typically non-natural) nucleotides that are incorporated during SMS, or by structurally modifying the polymerase to increase exonuclease activity. Random mutation strategies can also be used, e.g., in conjunction with appropriate screening steps to select libraries of mutants for increased branching (or other properties of interest). Combinations of random and site-directed mutagenesis can also be used, typically in conjunction with selection of mutant libraries for a property of interest.

[0062] As will be appreciated, polymerase enzymes of the present invention are not limited to DNA polymerases. The present invention provides modified recombinant reverse transcriptase enzymes and modified recombinant DNA-dependent RNA polymerases, which can exhibit increased branching fractions during RNA template sequencing and RNA synthesis, respectively.

[0063] One class of example mutants described in this application were designed to address branching fraction by modifying various sites in, e.g., a Φ29 polymerase, predominantly in and around the binding pocket, to create weaker polymerase-analog interactions during an extension (polymerization) reaction. A second class of example mutants described in this application were designed to increase branching by modifying various sites in, e.g., a Φ29 polymerase, predominantly in and around the exonuclease domain in order to increase the exonuclease rate of the polymerase to about 10% to 50% as compared to its polymerization rate. As noted, the "branching fraction" is the proportion of cognate nucleotide (or nucleotide analog, e.g., a dye-labeled analog) dissociation events from the polymerase active site to the total number of events, e.g., the sum of the incorporation events and dissociation events. For the purposes of this disclosure, dissociation events also include cleavage of an incorporated nucleotide as a result of exonuclease activity. These mutational features, i.e., increased branching by creating a more loosely structured binding pocket or increasing exonuclease activity, can be provided in combination.

[0064] Desirably, the branching fraction for a polymerase for a given nucleotide of interest (e.g., a labeled nucleotide analog) can be more than 50%, more preferably more than 60%, yet more preferably more than 70%, and still more preferably more than 80% or more of the total interactions, e.g., dissociation events and association events, of the nucleotide analog with the polymerase binding pocket. In comparison, a parental Φ29 polymerase exhibits a branching fraction of approximately 23% for, e.g., a thymine nucleotide analog that includes an Alexa568 fluorophore (Invitrogen Inc., Carlsbad, CA) linked to the terminal phosphate of a hexaphosphate chain, also referred to as A568dT6P, wherein approximately 23% of the total events with a gamma-linked A568dT6P nucleotide analog in the polymerase binding pocket are dissociation events.

[0065] The invention provides methods for generating recombinant polymerases that comprise modifications that increase the frequency of branching, which can be useful in any number of applications where accuracy of polymerization is beneficial, e.g., high-throughput sequencing systems, e.g., in a nanohole (an aperture of less than 1 μM diameter through which a synthesis complex can be illuminated by optical energy or monitored electrochemically) or specialized nanoholes such as zero-mode waveguides (ZMW), SNP genotyping using single base extension methods, real time monitoring of amplification, e.g., RT-PCR methods, and the like. Also provided by the invention are compositions that include such polymerases and methods in which these polymerases can be useful in, e.g., sequencing or making DNA.

[0066] In some embodiments, the compositions can also include nucleotide analogs, and preferably, optically labeled, e.g., fluorescently labeled, nucleotide analogs. In particularly preferred aspects, the compositions will include one or more types of phosphate-labeled nucleotide analog or analogs, e.g., a nucleotide analog comprising from 3-7 phosphate groups which in preferred cases may include a fluorophore coupled to the phosphate chain that is released upon incorporation, which can be incorporated into a copy nucleic acid by the modified polymerase in response to a DNA template. In some embodiments, the compositions can be present in a sequencing system, e.g. in a nanohole or specialized nanohole such as a zero-mode waveguide, where a polymerase of the invention can optionally be immobilized on a surface.

[0067] Modification of a polymerase, e.g., any of the polymerases described herein, or polymerases homologous to those described herein, by any one or more the strategies described herein can increase the frequency of branching events by creating a more loosely structured binding pocket for non-native nucleotides. The modified polymerases can comprise at least one amino acid substitution or a combination of amino acid substitutions relative to the parental polymerase.

[0068] Polymerases of the invention can be described or analyzed based upon comparison to a reference (e.g., parental) polymerase. For example, a parental polymerase (e.g., a wild type polymerase from which the polymerase of the invention is derived, or an available mutant) can serve as a reference polymerase. Comparisons between a reference polymerase and a polymerase of the invention are performed under selected reaction conditions. In general, standard reaction conditions can be defined based on the usual reaction conditions (e.g., optimized for the reaction at issue) for a given parental (e.g., wild-type) polymerase. That is, the reaction condition preferences for many polymerases are known; the appropriate adaptations of these conditions to specific applications such as SMS are known or can be determined through routine optimization or reaction conditions, and these "application optimized" reaction conditions can be used for comparison of the polymerase of the invention to the reference polymerase. For example, reaction conditions can be optimized for a reference polymerase such as a Φ29 polymerase, e.g., in an SMS application, with the reaction conditions being selected for optimal processivity, optimal fidelity, increased or decreased branch fraction, or a combination thereof, with that reference polymerase being used in a comparison to a polymerase of the invention under the optimized conditions. For example, in one SMS application, the reaction conditions can include those described above, and comparison to the polymerase of the invention can be conducted by performing a comparative assay, using the methods herein.

[0069] A number of specific examples of modified polymerases, e.g. modified to increase the average branching fraction, are described herein. The binding pocket is a portion of the polymerase that encompasses the nucleotide binding site and analog base during the pairing of a nucleotide analog with a template DNA. Because of the physical proximity of the binding pocket to the incoming nucleotide or nucleotide analog, mutations to this region can

affect the branching fraction. However, mutations that increase the branching fraction are not limited to this area of the polymerase. For example, relative to a parental or wild-type Φ29 DNA polymerase, useful modifications can include any of the following mutations: N62D and Y454A; D362S; Y259H; F237Y; L381I; Y369H; H461Y; A377G; K138Q; H461D; A377S; N62D and K371Q; V118L; and K124R. For the purposes of this disclosure, a parental polymerase includes a wild-type or available mutant/recombinant polymerase which is additionally altered to produce the desired properties of the invention, e.g., increased branching, delayed translocation or increased nucleotide analog residence time. A list of specific useful Φ29 mutants and their branching fractions (% BF) and corresponding increases in branching fraction (% increase BF) for a particular 6P nucleotide analog relative to a reference Φ29 polymerase (N62D) is provided in Table A below.

Table A

	A568dT6P		A647dG6P	
Mutation(s)	%BF	% Increase BF	%BF	% Increase BF
N62D	23	-	38	-
N62D_Y454A	30	30	48	25
D362S	31	34	46	20
Y259H	33	44	52	37
F237Y	35	53	46	21
L381I	37	60	47	22
Y369H	40	75	55	43
H461Y	43	86	57	49
A377G	44	90	75	97
K138Q	45	96	61	59
H461D	46	102	62	62
A377S	61	167	70	84
N62D_K371Q	62	172	94	145
V118L	65	181	74	92
K124R	73	216	84	119
	A555dC6P		A660dA6P	
Mutation(s)	%BF	% Increase BF	%BF	% Increase BF
N62D	27	-	22	-
N62D_Y454A	38	42	23	3
D362S	39	45	26	20
Y259H	39	44	25	11
F237Y	41	52	25	11
L381I	37	36	26	18
Y369H	43	59	28	27

	A555dC6P		A660dA6P	
Mutation(s)	%BF	% Increase BF	%BF	% Increase BF
H461Y	50	86	31	42
A377G	37	36	32	47
K138Q	48	78	27	21
H461D	51	91	31	39
A377S	72	168	56	152
N62D_K371Q	87	224	82	272
V118L	68	154	47	114
K124R	79	195	63	183

[0070] As noted, the branching fraction, e.g., % branching, is a relative measure of the number of times a correctly paired base, e.g., a Watson-Crick paired base, leaves the active site of the polymerase without forming a phosphodiester bond with the 3'OH of the primer-terminus relative to the total number of interactions that occur between a nucleotide (or nucleotide analog) and the binding pocket of the polymerase, e.g., the total number of opportunities the nucleotide or nucleotide analog has to correctly pair and incorporate. Additionally, for the purposes of this disclosure, branching refers to cleavage and dissociation from the polymerase active site of an incorporated nucleotide as a result of exonuclease activity. Branching is expressed as a percentage of the dissociation events vs. the total sum events, e.g., dissociation and association events. For example, for a polymerase harboring the Y369H mutation, for every 100 times an A568dT6P analog (i.e., a thymidine hexaphosphate nucleotide in which the terminal phosphate is labeled with an Alexa568 dye) interacts with the binding pocket of this polymerase, 40 of the events are non-productive dissociation events, e.g., wherein the analog dissociates from the polymerase instead of participating in a polymerization reaction. For this polymerase, the percent increase in branching fraction is 75% as compared to a reference phi29 polymerase (N62D) under identical reaction conditions.

[0071] The branching fraction can be measured by loading a polymerase active site with a cognate-matching nucleotide analog that can bind in the +1 and +2 positions. In the absence of divalent cation, this nucleotide cannot be incorporated into the DNA strand, so will pair with the template nucleotide at the +1 position but be released at some frequency specific for that analog/polymerase combination, e.g., the branching rate. This loading reaction is then followed by addition of a divalent cation that supports extension, e.g., Mn^{2+}, and a terminating–

type nucleotide analog, e.g., a dideoxynucleotide, comprising the same base as the cognate-matching analog in the loading step.

[0072] The dideoxy-analog will be incorporated into any +1 sites that are unoccupied and, once added, preclude further extension. Hence, polymerase active sites that are already occupied by a paired analog base extend to the +2 position, those that are not occupied (i.e. "branched") incorporate the dideoxy-type analog at +1 and do not extend, resulting in a single base addition. The extension products of this reaction are visualized by standard separation methods, e.g., gel or capillary electrophoresis, and the ratio of terminated product that is generated when a dideoxynucleotide is incorporated at the +1 position divided by the total terminated product, e.g., when a dideoxynucleotide is incorporated at both the +1 and +2 positions, indicates the fraction of 'branched' events that occur.

[0073] The branching fraction exhibited by modified polymerases of the present invention, e.g., a modified Φ29 polymerase, can be greater than a branching fraction exhibited by the corresponding wild-type polymerase for a given nucleotide analog. For example, a modified recombinant polymerase of the invention can exhibit an increased branching fraction that is greater than about 20% for the phosphate-labeled analog, greater than 50% for the phosphate-labeled analog, greater than 75% for the phosphate-labeled analog, greater than 100% for the phosphate-labeled analog, greater than 150% for the phosphate-labeled analog, or greater than 200% for the phosphate-labeled analog, as compared to the corresponding wild-type polymerase, e.g., a wild-type Φ29 polymerase, under the standard reaction conditions described above.

[0074] In some embodiments, the modified polymerase that exhibits an increased frequency of branching can also exhibit a K_m for a given phosphate-labeled nucleotide analog, e.g., any of the phosphate-labeled nucleotide analogs described herein, that is useful to achieve increased branching. For enzymes obeying simple Michaelis-Menten kinetics, kinetic parameters are readily derived from rates of catalysis measured at different substrate concentrations. The Michaelis-Menten equation, $V=V_{max}[S]([S]+K_m)^{-1}$ relates the concentration of uncombined substrate ([S], approximated by the total substrate concentration), the maximal rate (V_{max}, attained when the enzyme is saturated with substrate), and the

Michaelis constant (K_m, equal to the substrate concentration at which the reaction rate is half of its maximal value), to the reaction rate (V). To determine a K_m for a particular analog a series of extension reactions are performed with a varying concentration of the analog of interest with a fixed, saturating concentration of native nucleotides. A fit of the rate versus the substrate concentration generates estimation of the -K_m as the slope of this line.

[0075] The present invention also provides polymerases with increased exonuclease activity for increasing the branching fraction during a template-dependent polymerization reaction, e.g., SMS. In a preferred aspect, a polymerases of the invention exhibits an exonuclease rate that is between about 10% and 50% as compared to its polymerization rate.

[0076] As will be appreciated that the above-identified modified or recombinant polymerases that display increased branching fractions may optionally include additional modifications that confer other useful properties described herein, e.g., delayed translocation, increased nucleotide analog residence time and/or increased processivity. As will also be appreciated, the above-identified modified or recombinant polymerases that display increased branching fractions can be used in combination with any other embodiments described herein, including: (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) iterative sampling of unincorporatable nucleotides; (3) two slow-step enzyme systems; (4) detection of noncognate branching events; (5) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (6) modified recombinant polymerases that exhibit increased nucleotide residence time; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein.

C. Enhanced Sequencing Using Iterative Sampling of Unincorporatable Nucleotide Analogs

[0077] The present invention also employs nucleotide based competitive reagent compositions for identifying sequence elements, despite not being incorporated in a nascent nucleic acid strand. In particular, the unincorporatable nucleotide analogs of the invention, while not being incorporatable, may be nonetheless capable of specifically associating with the polymerase enzyme. That is, the polymerase will sample the unincorporatable nucleotides,

retaining them within the active site for a greater length of time than nucleotides that are not complementary to the position in the template nucleic acid, and release them when they cannot be incorporated. By providing different types of nucleotide or nucleoside analogs, e.g., mimetics of A, G, T, C, and/or U, bearing distinguishable labels, e.g., spectrally resolvable fluorophores or other labeling groups, one can monitor the sampling of these nucleotides as an indication of the nucleotide that is next to be incorporated. For example, one may provide labeled, unincorporatable nucleotide analogs at concentrations in excess of incorporatable nucleotides, e.g., 2X, 5X or even 10X or greater. Each incorporation of an incorporatable nucleotide will, by virtue of the excess concentration, be preceded by repeated sampling events of the unincorporatable nucleotides, which will each carry its associated signal event. The incorporatable nucleotides may then either bear no label, or preferably, bear a label that is distinguishable from the unincorporatable nucleotides, so as to mark the termination of the sampling of a given base and proceeding onto the next base in the sequence. In such cases, it may be desirable to label all incorporatable nucleotides with a single type of fluorophore, i.e., indistinguishable from the label groups on the other types of incorporatable nucleotides present, but distinguishable from all of the unincorporatable nucleotides.

[0078] The signal detection for the foregoing process is schematically illustrated in **Figure 4**. In particular, Figure 4 shows a schematic illustration of a set of signal traces from a single molecule sequence by incorporation reaction. As shown, the plot shows five signal traces, one for each type of differentially labeled unincorporatable nucleotide analog (indicated as A', T', G' and C', as well as a trace for the signal associated with the type of label coupled to the incorporatable nucleotide (labeled as "I"). As shown, repeated sampling of the cognate unincorporatable nucleotide analog, e.g., A', provides an iterative set of signal events 402, followed by a signal 404 on the I trace indicating conclusion of the incorporation event. This pattern is repeated for the next base to be incorporated (indicated by iterative signals 406 in the T' trace, followed again by the incorporation signal 408, in the I trace, and again by the iterative sampling signal 410 in the A' trace followed by the incorporation signal 412 in the I trace. Because these unincorporatable nucleotides are mimetic of the base to be incorporated, they possess a longer retention time in the active site than the analog that is not complementary to the next base in the template, and as such, provide a signal profile that is distinguishable

from random, incorrect sampling, e.g., as indicated by transient signal events 414. Such iterative sampling may include two, three, four, five, ten or greater than ten signal events for each incorporation.

[0079] As noted above, the competitive reagents used are going to be non-reactive in the reaction of interest. In preferred aspects, and without being bound to any particular theory of operation, the competitive compounds may possess structures similar to nucleotides or portions thereof, such that they can competitively interact with the reaction of interest, e.g., through association with the polymerase active site. By way of example, such structures may comprise a polyphosphate component, e.g., a pyrophosphate, triphosphate, tetraphosphate, pentaphosphate, or longer phosphate chain, so that the compound mimics one or more of a nucleotide or the product of a polymerase mediated incorporation reaction, which is capable of competitively interacting with the polymerase, relative to the nucleotide analogs.

[0080] In certain preferred cases, additional components may be coupled to the polyphosphate component that mimic other portions of the nucleotide or nucleotide analog. By way of example, the polyphosphate component may be coupled to a cyclic and/or aromatic component that may structurally mimic the nucleoside component in its interaction with the polymerase. Such structures are generally illustrated by the following structure:

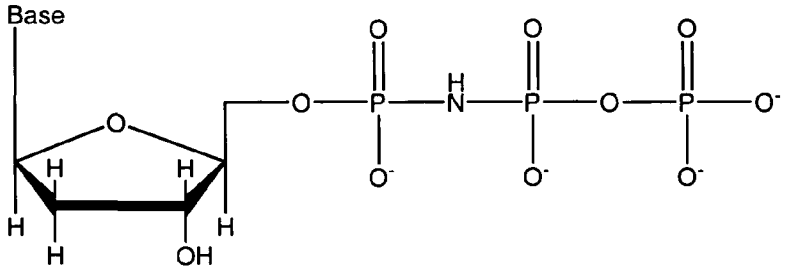
$$P\text{-}(P)_n\text{-}A;$$

where P is a phosphate or phosphonate group, n is an integer from 1 to about 6, and A includes a cycloalkyl or aryl group, a carbohydrate group, or the like.

[0081] In the case of nucleotide analogs used in analytical primer extension reactions, e.g., in nucleic acid sequence analysis, such nucleotide analogs will be unincorporatable in such primer extension reaction by the polymerase used. Further, in preferred aspects, such unincorporatable analogs will typically still be capable of interaction with the polymerase, e.g., active site binding, but will be unable to be incorporated in a primer extension reaction. In preferred aspects, this is accomplished by providing nucleotide analogs that possess nonhydrolyzable groups within the phosphate chain, such that the phosphoester linkage between the analog and the primer strand, cannot be formed, as mediated by the polymerase. One particularly effective approach to producing an unincorporatable nucleotide analog

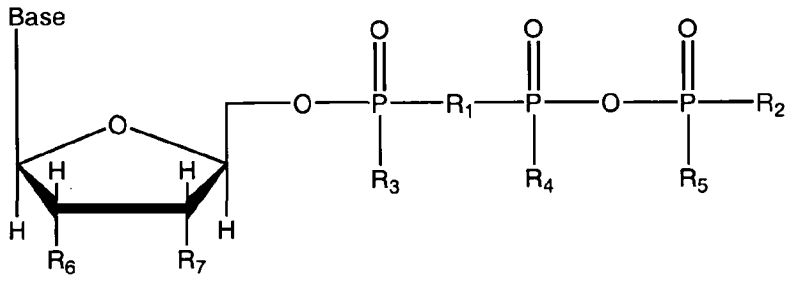
includes replacing the phosphoester linkage between the alpha and beta phosphate of a nucleoside polyphosphate with a nonhydrolyzable linkage.

[0082] One example of such an analog is illustrated below, where the oxygen group between the alpha and beta phosphate groups is replaced with an nonhydrolyzable linkage, such as the illustrated amino group.



[0083] Although illustrated as an amino linkage, it will be appreciated that a variety of other linkages may be used between the alpha and beta phosphates, e.g., an amino, methyl, thio, or other linkages not hydrolyzed by polymerase activity. Additionally, although illustrated as including three phosphate groups analogous to a nucleoside triphosphate, it will be appreciated that other polyphosphate configurations may be employed in the invention, including, for example, tetraphosphate analogs, pentaphosphate analogs, hexaphosphate analogs, and the like.

[0084] Thus, the structures employed in certain preferred aspects of the invention may generally be described with reference to the following structure:



where R_1 comprises a linking group that is non-hydrolyzable by the polymerase enzyme being used. Particularly preferred linkages include amino linkages, alkyl linkages, e.g., methyl, and thio linkages. While R_2 may comprise oxygen, in some preferred aspects, it will include additional phosphate groups, e.g., mono-, di-, or triphosphate groups coupled to the gamma

phosphate group. Alternatively or additionally, the R_2 group may include, in addition to or instead of additional phosphate groups, labeling functionalities that provide for the detection of the competitive substrates, but still permit its distinguishing from the incorporatable nucleotides. In other aspects, the R_2 group may include moieties that provide other functionalities to the reaction system other than as a labeling group. For example, R_2 may comprise an agent that reduces the potential for photodamaging effects on a polymerase enzyme, either coupled directly to the terminal phosphate group, or through a linking group.

[0085] The relative concentration of the competitive substrates to the incorporatable substrates, within a reaction mixture may generally be varied in accordance with a desired application. In particular, because the concentration of the competitive substrates affects the interactions of the complex with the incorporatable nucleotides, one can modulate those interactions by altering the ratios between incorporatable nucleotides and competitive substrates. In typical applications, however, the relative molar concentration of competitive substrate will range from about 0.5X to about 10X, 20X or greater of the concentration of the actual substrates (or incorporatable nucleotide analogs). Thus, the concentration ratio of unincorporatable nucleotide analogs to incorporatable nucleotide analogs will typically range from a lower ratio of from about 0.1:1, 0.2:1, 0.5:1 and 1:1, to an upper ratio of about 2:1, 3:1, 5:1, 10:1 or even 20:1, with each iteration of the foregoing being encompassed in the disclosure hereof.

[0086] As will be appreciated, iterative sampling of unincorporatable nucleotides can be used in combination with any other embodiments described herein, including: (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (3) two slow-step enzyme systems; (4) detection of noncognate branching events; (5) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (6) modified recombinant polymerases that exhibit increased nucleotide residence time; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein.

D. Enhanced Sequencing Using Two Slow-Step Enzyme Systems Combined With Detection and Analysis of Branch Fraction Nonincorporation Events

[0087] Enzyme systems that exhibit kinetic mechanisms having two or more slow, kinetically observable, or partially rate-limiting reaction steps within an observable phase of the polymerase reaction can be useful for example, in single-molecule, real-time observations of such enzyme activity, which rely, at least in part, on detecting and identifying the enzyme reaction as it is occurring. By designing the reaction system to have two or more partially rate-limiting steps (i.e., "two slow-step" enzyme systems), the relative number of short, difficult to detect, events can be lowered. Details regarding enzyme systems exhibit kinetic mechanisms having two or more slow, kinetically observable, or partially rate-limiting reaction steps within an observable phase of the polymerase reaction can be found in Bjornson et al. PCT Application Serial Number PCT/US2009/002003 TWO SLOW-STEP POLYMERASE ENZYME SYSTEMS AND METHODS, incorporated herein by reference in its entirety for all purposes.

[0088] Certain types of template nucleic acid sequences present unique challenges during single molecule sequencing. For example, during single molecule sequencing of homonucleotide stretches (i.e., a portion of a template nucleic acid having two or more consecutive bases that are identical), ascertaining the number of nucleotide incorporation events that are represented by a series of signal pulses generated under high branch fraction sequencing conditions can be difficult using previous typical enzyme systems that exhibit one slow-step. By way of example, if 20 consecutive signal pulses are detected from a fluorescent-labeled nucleotide analog of dATP (e.g., A488dA4P, see below), a system that detects signal pulses and analyzes the time-sequence of those signal pulses for purposes of calling bases may not be able to accurately determine how many dATP analog incorporation events (i.e., the number of consecutive T residues in the template sequence) are represented by the 20 signal pulses. Accordingly, enzyme systems that permit a more predictable distribution of nucleotide binding events per incorporation are desirable, because, e.g., the number of nucleotides within a homonucleotide stretch can be determined by using a multiple of the expected number of binding events per incorporation.

[0089] In one aspect, the present invention provides two slow-step enzyme systems that provide a more predictable frequency or rate of nucleotide binding events (e.g., the number of times a nucleotide or nucleotide analog samples the active site of the polymerase) per incorporation event. While not being bound by any particular theory, the following theoretical basis is provided for obtaining improved single-molecule sequencing results, e.g., for homonucleotide stretches, by using a system having two or more slow steps within an observable phase. A model for the effect of two slow steps on the number of nucleotide binding events is described herein and illustrated in **Figure 5**. Figure 5 shows a plot of calculated number of binding events per incorporation for cases in which (1) one step is rate-limiting and (2) two equivalent partially rate-limiting (slow) steps are present for the observable phase in which the nucleotide is associated with the enzyme before unbinding.

[0090] For the case in which one step is rate-limiting, the distribution for the number of binding events per incorporation can be represented by the single exponential equation:

$$y = A_0 e^{-kt} \qquad \text{Eq. 1}$$

This represents the case in which, for example, one conformational change of the enzyme after nucleotide binding is the single slow step.

[0091] Figure 5 illustrates that where one slow-step is present in this phase, there is an exponentially decreasing number of binding events per incorporation, providing a distribution in which there is a relatively high probability that the number of binding events per incorporation will be low. In this scenario, it can be more challenging to distinguish homonucleotide stretches of the same base in the DNA template.

[0092] For the case in which there are two slow steps associated with binding of a nucleotide, for example via two consecutive conformational changes with similar rate constants, the number of binding events can be represented by a sum of two exponentials with an equation:

$$y = A_0 e^{-k_1 t} - B_0 e^{-k_2 t} \qquad \text{Eq. 2}$$

[0093] Figure 5 illustrates that for the case in which there are two slow steps, the probability of one or a low number of binding events per incorporation is relatively low as compared to enzyme systems having one slow step. In addition, the probability distribution for two slow steps exhibits a peak in the plot, with the most frequently observed number of binding events per incorporation greater than zero. This type of distribution can be advantageous for single-molecule sequencing where it is desired to resolve homonucleotide stretches. In this scenario, the expected number of binding events per incorporation will be a corresponding multiple factor of the most frequent occurrence of binding events per incorporation as governed by this distribution.

[0094] The two slow steps can include, e.g., nucleotide addition, enzymatic isomerization (such as to or from a closed state), and cofactor binding or release. The use of a distribution of pulses to determine a kinetic mechanism having two slow (kinetically observable) steps is described, for example, in Miyake et al. Analytical Chemistry 2008 80 (15), 6018-6022. The determination of the steps in a multistep reaction such as a polymerase reaction is described, for example, in Zhou, et al. J. Phys. Chem. B, 2007, 111, 13600-13610.

[0095] As noted above, the present invention provides enzyme systems that exhibit kinetic mechanisms having two or more slow, kinetically observable, or partially rate-limiting reaction steps within an observable phase of the polymerase reaction observable phase will generally have a time period during which the phase is observable. The time period for a bright phase, for example, can be represented by the pulse width. The time period for a dark phase can be represented, for example, by the interpulse distance. The length of each time period will not be the same for each nucleotide addition, resulting in a distribution of the length of the time periods. In some cases, the time periods with the shortest length will not be detected, leading to errors, for example in single-molecule sequencing. We have found that by designing enzyme systems such as polymerase reaction systems in which there are two slow, or kinetically observable, steps within an observable phase, the relative number of short, unobservable, time periods can be reduced, resulting in a higher proportion of observable sequencing events, and allowing for a more accurate determination of nucleotide sequence. As used herein, an observable phase includes phases that are not directly observable, but can be ascertained by measurements of other, related phases. For example, the lengths of dark phases can be

observed by measuring the times between optical pulses corresponding to a related bright optical phase. Also as described herein, a phase which is dark under some labeling conditions can be bright under other labeling conditions.

[0096] In natural polymerase-mediated nucleic acid synthesis, a complex is formed between a polymerase enzyme, a template nucleic acid sequence, and a priming sequence that serves as the point of initiation of the synthetic process. During synthesis, the polymerase samples nucleotide monomers from the reaction mix to determine their complementarity to the next base in the template sequence. When the sampled base is complementary to the next base, it is incorporated into the growing nascent strand. This process continues along the length of the template sequence to effectively duplicate that template. Although described in a simplified schematic fashion, the actual biochemical process of incorporation is relatively complex.

[0097] The process can be described as a sequence of steps, wherein each step can be characterized as having a particular forward and reverse reaction rate that can be represented by a rate constant. One representation of the incorporation biochemistry is provided in **Figure 6**. It is to be understood that the scheme shown in Figure 6 does not provide a unique representation of the process. In some cases, the process can be described using fewer steps. For example, the process is sometimes represented without inclusion of the enzyme isomerization steps 606 and 610. Alternatively, the process can be represented by including additional steps such as cofactor binding. Generally, steps which can be slow, and thus limit the rate of reaction will tend to be included. The present invention relates to methods, systems, and compositions in which the polymerization reaction has two or more slow steps within certain phases of the polymerase reaction. Various schemes can be used to represent a reaction having two slow steps that may have more or fewer identified steps. In some cases the two or more slow steps are consecutive. In some cases, there can be intervening fast steps between the two or more slow steps.

[0098] As shown in Figure 6, the synthesis process begins with the binding of the primed nucleic acid template (D) to the polymerase (P) at step 602. Nucleotide (N) binding with the complex occurs at step 604. Step 606 represents the isomerization of the polymerase from the open to closed configuration. Step 608 is the chemistry step where the nucleotide is

incorporated into the growing strand of the nucleic acid being synthesized. At step 610, polymerase isomerization occurs from the closed to the open position. The polyphosphate component that is cleaved upon incorporation is released from the complex at step 612. The polymerase then translocates on the template at step 614. As shown, the various steps can include reversible paths and may be characterized by the reaction constants shown in Figure 6 where:

[0099] k_{on}/k_{off}=DNA binding/release;

[0100] k_1/k_{-1}= nucleotide binding/release;

[0101] k_2/k_{-2}=polymerase isomerization (open/closed);

[0102] k_3/k_{-3}=nucleotide incorporation (chemistry);

[0103] k_4/k_{-4}=polymerase isomerization (closed/open);

[0104] k_5/k_{-5}=polyphosphate release/binding;

[0105] k_6/k_{-6}=polymerase translocation.

[0106] Thus, during steps 604 through 610, the nucleotide is retained within the overall complex, and during steps 604 and 606, reversal of the reaction step will yield an unproductive event, i.e., not resulting in incorporation. For example, a bound nucleotide at step 604 may be released regardless of whether it is the correct nucleotide for incorporation.

[0107] By selecting the appropriate polymerase enzyme, polymerase reaction conditions, and polymerase substrates, the absolute and relative rates of the various steps can be controlled. We have found that controlling the reaction such that the reaction exhibits two or more kinetically observable, or slow steps can produce a nucleic acid polymerization reaction in which the incorporation of the nucleotides can be observed more accurately. These characteristics are particularly useful for sequencing applications, and in particular single-molecule DNA sequencing.

[0108] In some cases, the invention involves a process having two or more kinetically observable steps that comprise steps after nucleotide binding through the step of product release. For the mechanism shown in Figure 6, this would be, for example, any of steps 606,

608, 610, and 612. In some cases, steps 608 (nucleotide incorporation) and 612 (product release) are the two slow, or kinetically observable steps. As noted previously, where one desires systems with slow steps in a dark phase, the invention may involve a process having two or more slow steps that comprise the steps after product release through nucleotide binding. For the mechanism shown in Figure 6, this would include steps 614 and 604.

[0109] In some cases, the invention involves a process in which there are two or more slow steps in two different observable phases within the polymerization, for example, two slow steps in a bright phase and two slow steps in a dark phase. For example, this could include a system having two slow steps in the steps after nucleotide binding through product release, and two slow steps for the steps after product release through nucleotide binding.

[0110] As is described herein, producing a process in which there are two or more slow steps in these portions of the polymerase reaction can result in a higher proportion of detectable enzyme states which can be useful, for example, to observe the sequential incorporation of nucleotides for nucleotide sequencing.

[0111] By the term slow-step we generally mean a kinetically observable step or partially rate-limiting step. The slow step need not be slow in the absolute sense, but will be relatively slow as compared with other steps in the enzymatic reaction. The slow, or kinetically observable steps, can be, for example, each partially rate-limiting, in that the rate of the step has a measurable effect on the kinetics of the enzymatic reaction. An enzymatic process, such as nucleic acid polymerization, can have both slower, kinetically observable steps and faster steps which can be so fast that they have no measurable effect on the kinetics, or rate, of the reaction. In some reactions, there can be a single rate-limiting step. For such reactions, the kinetics can be characterized by the rate of that single step. Other reactions will not have a single rate-limiting step, but will have two or more steps which are close enough in rate such that the characteristics of each will contribute to the kinetics of the reaction. A kinetically observable step is generally a step which is slow enough relative to the other steps in the reaction such that it can be experimentally ascertained. The experimental identification of a kinetically observable step can be done by the methods described herein, or by methods for assessing the kinetics of chemical and enzymatic reactions known in the art. For the current invention, the

slow, or kinetically observable steps, need not be the slowest step or the rate-limiting step of the reaction. For example, a process of the current invention can involve a reaction in which step 604, nucleotide addition is the slowest (rate-limiting) step, while two or more of steps 606, 608, 610, or 612 are each kinetically observable.

[0112] As used herein, the term rate, as applied to the steps of a reaction can refer to the average rate of reaction. For example, when observing a single-molecule reaction, there will generally be variations in the rates as each individual nucleotide is added to a growing nucleic acid. In such cases the rate of the reaction can be represented by observing a number of individual events, and combining the rates, for example, by obtaining an average of the rates.

[0113] As used herein, the reference to the rate of a step or rate constant for a step can refer to the forward reaction rate of the polymerase reaction. As is generally understood in the art, reaction steps can be characterized as having forward and reverse rate constants. For example, for step 608, k_3 represents the forward rate constant, and k_{-3} represents the reverse rate constant for the nucleotide incorporation. Some reaction steps, such as step 608, constitute steps which would be expected to be first order steps. Other steps, such as the forward reaction of step 604, with rate constant k_2, would be expected to be second order rate constants. For the purposes of the invention, for comparing the rate or the rate constant of a first order to a second order step, the second order rate constant k_2 can be treated as a pseudo-first order rate constant with the value [N]*k_2 where the concentration of nucleotide [N] is known.

[0114] It is generally desirable that the kinetically observable steps of the invention have rate constants that are lower than about 1000 per second. In some cases, the rate constants are lower than about 500 per second, lower than about 200 per second, lower than about 100 per second, lower than about 60 per second, lower than about 50 per second, lower than about 30 per second, lower than about 20 per second, lower than about 10 per second, lower than about 5 per second, lower than about 2 per second, or lower than about 1 per second.

[0115] In some embodiments the slowest of the two or more kinetically observable steps has a rate constant when measured under single-molecule conditions of between about 500 to about 0.1 per second, about 200 to about 0.1 per second, about 60 to about 0.5 per second, about 30 per second to about 2 per second, or about 10 to about 3 per second.

[0116] The ratio of the rate constants of each the two or more slow steps is generally greater than 1:10, in some cases the ratio of the rate constants is about 1:5, in some cases the ratio of the rate constants is about 1:2, in some cases, the ratio of rate constants is about 1:1. The ratio of the rate constants can be between about 1:10 and about 1:1, between about 1:5 and about 1:1, or between about 1:2 and about 1:1.

[0117] In some cases it is useful to consider the two slow-step system in terms of rates rather than rate constants. It is generally desirable that the kinetically observable steps of the invention have rates that are lower than about 1000 molecules per second when the reactions are carried out under single-molecule conditions. In some cases, the rates are lower than about 500 molecules per second, lower than about 200 molecules per second, lower than about 100 molecules per second, lower than about 60 molecules per second, lower than about 50 molecules per second, lower than about 30 molecules per second, lower than about 20 molecules per second, lower than about 10 molecules per second, lower than about 5 molecules per second, lower than about 2 molecules per second, or lower than about 1 molecule per second.

[0118] In some embodiments the slowest of the two or more kinetically observable steps has a rate when measured under single-molecule conditions of between about 500 to about 0.01 molecules per second, between about 200 to about 0.1 molecules per second, between about 60 to about 0.5 molecules per second, about 30 molecules per second to about 2 molecules per second, or about 10 to about 3 molecules per second.

[0119] The ratio of the rates of each the two or more slow steps is generally greater than 1:10, in some cases the ratio of the rates is about 1:5, in some cases the ratio of the rates is about 1:2, in some cases, the ratio of rates is about 1:1. The ratio can be between about 1:10 and about 1:1, between about 1:5 and about 1:1, or between about 1:2 and about 1:1.

[0120] A two or more slow-step system of the present invention can be obtained by selecting the correct set of polymerase enzyme, polymerase reaction conditions, and polymerase reaction substrates.

[0121] Table B presents exemplary Φ29 mutants that can exhibit two slow step behavior under appropriate reaction conditions. The first three modified polymerases exhibit

the most pronounced two slow step behavior, followed by the next six. As noted, the polymerases are optionally exonuclease-deficient; for example, they can also include an N62D substitution.

Table B

A484E/E375Y/K512Y/T368F
A484Y/E375Y/K512Y/T368F
N387L/E375Y/K512Y/T368F
T372Q/E375Y/K512Y/T368F
T372L/E375Y/K512Y/T368F
T372Y/K478Y/E375Y/K512Y/T368F
I370W/E375Y/K512Y/T368F
F198W/E375Y/K512Y/T368F
L381A/E375Y/K512Y/T368F
E375Y/K512Y/T368F

[0122] The polymerase reaction conditions can also be important for obtaining a two slow-step enzyme system. In particular, polymerase reaction conditions include components selected to produce two slow-step kinetics. The polymerase reaction conditions include the type and concentration of buffer, the pH of the reaction, the temperature, the type and concentration of salts, the presence of particular additives which influence the kinetics of the enzyme, and the type, concentration, and relative amounts of various cofactors, including metal cofactors. The term "polymerase reaction conditions" as used herein generally excludes the concentration of the polymerase enzyme or the concentration of the primer-template complex. Thus, two reactions are run under substantially the same polymerase reaction conditions where the first reaction has a small amount of polymerase enzyme, such as a single polymerase enzyme, and a small amount of primer template complex, such as a single primer-template complex associated with a single polymerase enzyme, and the second reaction has a higher concentration of polymerase enzyme, for example a concentration of polymerase enzyme of about 0.05 μM to 0.5 μM, and about 0.01 μM to about 0.1 μM.

[0123] It some embodiments the type and concentration of buffer are chosen in order to produce a reaction having two slow steps. Enzymatic reactions are often run in the presence of a buffer, which is used, in part, to control the pH of the reaction mixture. We have found that

in some cases the type of buffer can influence the kinetics of the polymerase reaction in a way that can lead to two slow-step kinetics. For example, in some cases, we have found that the use of TRIS as buffer is useful for obtaining a two slow-step reaction. Buffers suitable for the invention include, for example, TAPS (3-{[tris(hydroxymethyl)methyl]amino}propanesulfonic acid), Bicine (N,N-bis(2-hydroxyethyl)glycine), TRIS (tris(hydroxymethyl)methylamine), ACES (N-(2-Acetamido)-2-aminoethanesulfonic acid), Tricine (N-tris(hydroxymethyl)methylglycine), HEPES 4-2-hydroxyethyl-1-piperazineethanesulfonic acid), TES (2-{[tris(hydroxymethyl)methyl]amino}ethanesulfonic acid), MOPS (3-(N-morpholino)propanesulfonic acid), PIPES (piperazine-N,N'-bis(2-ethanesulfonic acid)), and MES (2-(N-morpholino)ethanesulfonic acid).

[0124] The pH of the reaction can influence the kinetics of the polymerase reaction, and can be used as one of the polymerase reaction conditions to obtain a reaction exhibiting two slow-step kinetics. The pH can be adjusted to a value that produces a two slow-step reaction mechanism. The pH is generally between about 6 and about 9. In some cases, the pH is between about 6.5 and about 8.0. For example, the pH can be by way of illustrations between about 6.5 and 7.5. The pH can be about 6.5, 6.6, 6.7, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4, or 7.5.

[0125] The temperature of the reaction can be adjusted in order to obtain a reaction exhibiting two slow-step kinetics. The reaction temperature may depend upon the type of polymerase which is employed. Temperatures between about 15^{O} C and 90^{O} C, between about 20^{O} C and 50^{O} C, between about 20^{O} C and 40^{O} C, or between about 20^{O} C and 30^{O} C can be used.

[0126] In one aspect, the present invention is directed to the use of a mixture of catalytic and non-catalytic metal ions in a nucleic acid synthesis reaction, to modulate the reaction kinetics of the complex. Thus, in at least one aspect, the invention is directed to nucleic acid synthesis reaction mixtures that include both catalytic and non-catalytic metals. The molar ratio of catalytic to non-catalytic metals in the reaction mixture will generally vary depending upon the type of kinetic modulation desired for a given synthesis reaction, where slower incorporation would suggest higher levels of non-catalytic metal ions. Typically, such ratios of catalytic to non-catalytic metals in the reaction mixture will vary from about 10:1 to

about 1:10, and preferably, from about 10:1 to about 1:5, depending upon the desired level of modulation, the particular enzyme system employed, the catalytic and non-catalytic metal cofactors that are used, and the reaction conditions. In particularly preferred aspects, the ratios of catalytic to non-catalytic metals will be in the range of from about 5:1 to about 1:1, with ratios of from about 2.5:1 to about 1.5:1 being particularly preferred.

[0127] In addition to the presence of such metals at the ratios described herein, the absolute concentration of such metals in the reaction mixtures will typically range from about 0.05 mM to about 50 mM, in some cases from about 0.1 mM to about 10 mM, in some cases from about 0.1 mM to about 5 mM. The composition can include, for example, from about 0.1 mM $MnCl_2$ to about 1 mM $MnCl_2$ and from about 0.1 mM $CaCl_2$ to about 2 mM $CaCl_2$; or from about 0.2 mM $MnCl_2$ to about 1 mM $MnCl_2$ and from about 0.4 mM $CaCl_2$ to about 1.5 mM $CaCl_2$.

[0128] As will be appreciated, the two slow-step enzyme systems described above can be used in combination with any other embodiments described herein, including: (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (3) iterative sampling of unincorporatable nucleotides; (4) detection of noncognate branching events; (5) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (6) modified recombinant polymerases that exhibit increased nucleotide residence time; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein.

E. Enhanced Sequencing By Detection of Noncognate Branching Events

[0129] Nucleic acid sequencing approaches that utilize fluorescent-labeled nucleotide analogs typically require the detection of at least four colors – one color for each nucleotide representing the four different bases A, T/U, C, or G. Sequencing by incorporation methods, e.g., real time single molecule sequencing, also typically involves the detection of at least four different fluorescent labels corresponding to each of four nucleotide analogs. New approaches for reducing the number of different fluorescent labels to be detected by a sequencing system

are provided herein; these approaches can reduce the cost of nucleic acid sequencing by reducing, e.g., the number of color channels and other associated optical capabilities of the detection system.

[0130] The present invention provides new methods for nucleic acid sequencing that reduce the number of required colors to be detected by the sequencing system. Such methods utilize noncognate branching to determine which nucleotide is incorporated at a particular incorporation site during a nucleic acid polymerization reaction. For the purposes of this disclosure, noncognate branching refers to the branching (or iterative sampling) of a nucleotide, where the base of the nucleotide would incorrectly pair (not form a correct Watson-Crick base pair) with the base at the incorporation site of the template nucleic acid. For example, when the incorporation site of a template nucleic acid contains the base guanine (or G), it has been observed that branching (or sampling) of dGTP nucleotides or nucleotide analogs occurs at an appreciable frequency. The frequency of noncognate branching is sufficient to permit the identification of the nucleotide actually incorporated, regardless of whether the nucleotide actually incorporated is labeled. For instance, multiple signal pulses derived from noncognate branching (i.e., iterative sampling) of fluorescently-labeled G-containing nucleotides or nucleotide analogs can be utilized by a signal detection and sequencing system to determine that a C-containing nucleotide (which does not require a fluorescent label) was actually incorporated at that site. Thus, the temporal sequence of incorporation of four different nucleotides can be ascertained by utilizing, at most, three different fluorescent labels.

[0131] Noncognate branching of T-containing nucleotides has also been observed when the base at the incorporation site of the template is T. Accordingly, it is possible to determine the sequence of a nucleic acid template using a two-color detection system, where noncognate branching of G- and T-containing nucleotides – alone – permits the identification of incorporation of C- and A-containing nucleotides, thereby eliminating the requirement that C- and A-containing nucleotides be labeled. As will be appreciated, the present invention provides enhanced sequencing using noncognate branching, where the sampling noncognate nucleotide can be any nucleotide that would incorrectly pair with the corresponding base of the template nucleic acid.

[0132] It will be appreciated that noncognate branching events, in conjunction with a sequencing system that detects and accounts for noncognate branching events, are useful for sequencing applications other than those designed to reduce, e.g., the number of color channels required of the sequencing system. For example, even where sequencing is performed using four colors, noncognate branching can be used to validate or assist in the determination by a sequencing system of which nucleotide was incorporated at a particular site. Thus, when the base at the incorporation site of a template of unknown sequence is G, and the quantity or quality of signal pulses generated from sampling of dCTP analogs does not unambiguously permit assignment of dCTP as the incorporating nucleotide at that site, noncognate branching of labeled dGTP nucleotides facilitates the determination that dCTP was indeed incorporated at the site.

[0133] As will be appreciated, enhanced sequencing by detection and analysis of noncognate branching events can be used in combination with any other embodiments described herein, including: (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (3) iterative sampling of unincorporatable nucleotides; (4) two slow-step enzyme systems; (5) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (6) modified recombinant polymerases that exhibit increased nucleotide residence time; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein.

II. SEQUENCING BY INCORPORATION USING MORE THAN ONE MODE OF SEQUENCING

[0134] A sequencing composition may be especially advantageous for one particular purpose or type of template nucleic acid, but may exhibit limitations for a second purpose or type of template nucleic acid. As described in detail herein, sequencing under conditions that promote a relatively high branching fraction, in combination with appropriate signal analysis, reduces the likelihood of base calling errors that result from undetected signal pulses generated from genuine nucleotide incorporation events. Accordingly, the base calling accuracy of

sequencing by incorporation reactions that exhibit a relatively high level of branch fraction nonincorporation events – when combined with a sequencing system that accounts for such redundant signal pulses per incorporation event – can be higher than the level of accuracy achieved under low branch rate conditions. On the other hand, sequencing under conditions in which the branching fraction is low typically results in improved sequence read length, as the number of incorporation events will be greater before the polymerase has an opportunity to dissociate from the template nucleic acid. Longer read lengths simplify assembly of contig information, e.g., to facilitate genomic sequencing.

[0135] De novo sequencing (sequencing a template nucleic acid of unknown sequence, e.g., genomic DNA of unknown sequence) is, in some cases, optimally performed under a combination of both high accuracy and high read length conditions. Identifying one particular sequencing composition that adequately fulfills both of these competing requirements is not necessary in the present invention. Instead, the present invention provides new methods and systems for sequencing a template nucleic acid, in which the template nucleic acid is sequenced using more than one sequencing mode ("multi-modal" or "variable mode" sequencing). As used herein, a sequencing mode refers to a sequencing composition (e.g., a mixture of a particular nucleic acid polymerase, nucleotides or nucleotide analogs, metal cofactors, and other components of a sequencing reaction) and other conditions that affect nucleic acid polymerization, e.g., reaction temperature. The methods and systems of the invention utilize a first sequencing mode that confers a particular benefit for purposes of generating a sequence readout (e.g., high accuracy), and then switch to a second (or more) sequencing mode that confers a benefit not realized during the first sequencing mode, e.g., a high sequence read length. As will be appreciated, any number of different modes can be employed by the methods and systems of the inventions, e.g., 2, 3, 4, 5, 10, or 20 different modes may be used to sequence a particular nucleic acid template until the desired results are achieved. For example, a template nucleic acid can be sequenced multiple times using a "high accuracy" mode and/or multiple times using a "high read length" mode, in order to obtain the accuracy and fold coverage useful for, e.g., de novo sequencing. Alternatively, a single template nucleic acid can be partially sequenced using, e.g., a high accuracy mode, and the mode can be switched prior to complete sequencing of the template to, e.g., a high read length mode.

[0136] A sequencing mode in which, e.g., high accuracy is desirable can employ, e.g., reaction conditions that induce a relatively high average branching fraction during the nucleic acid polymerization reaction. Such reaction conditions are described in detail herein. In one embodiment of the invention, one or more metal cofactors may be included in the sequencing composition at concentrations and/or ratios that induce a relatively high average branching fraction. Metal cofactors of the invention include, e.g., Mg^{++}, Mn^{++}, Zn^{++}, Co^{++}, Ca^{++}, Fe^{++}, Cr^{++}, and/or Sr^{++}. Reaction conditions in which metal cofactor identities, concentrations and/or ratios induce high levels of branching are described in detail herein. As described herein, a relatively high level of branching fraction nonincorporation events can be induced in a sequencing composition that includes, e.g., a relatively low concentration of Mn^{++}, a sequencing composition that includes Mg^{++}, a sequencing composition that includes both Mg^{++} and Mn^{++} with Mg^{++} being included at a higher concentration than Mn^{++}, and a sequencing composition that includes Mn^{++} and a metal cofactor other than Mg^{++}, e.g., calcium, barium, strontium, iron, cobalt, nickel, tin, zinc, and europium.

[0137] A sequencing mode in which, e.g., high read length is desirable can employ, e.g., reaction conditions that discourage branch fraction nonincorporation events during the nucleic acid polymerization reaction. Such reaction conditions are described herein. As described herein, reaction conditions that do not promote branching events include, e.g., Mn^{++} alone at a concentration at which branching is not induced (e.g., greater than 250 μM), or Mn^{++} in combination with a second metal cofactor, e.g., Mg^{++}, with the concentration of Mn^{++} being greater than the concentration of Mg^{++}. Desirably, sequencing modes of the present invention that produce high read lengths (e.g., a high number of successive incorporation events that are detected by sequencing systems of the invention) can produce read lengths of preferably more than 200 base pairs (bp) or more, more preferably 500 bp or more, more preferably 1000 bp or more, more preferably 10,000 bp or more, or more preferably 50,000 bp or more.

[0138] In order to permit the switching between a first sequencing mode and a second (or subsequent or more) sequencing mode, the sequencing composition – including the template nucleic acid and polymerase – is desirably confined within a structure to which additional components can be added to the sequencing composition to achieve the desired subsequent mode. In one aspect, for example, the sequencing of a template nucleic acid under

a first sequencing mode occurs within a structure to which a channel, e.g., a microfluidic channel, delivers the reagents necessary to achieve the subsequent mode. For example, switching from a high branch fraction sequencing mode, in which Mg^{++} is the sole catalytic metal ion, to a high read length sequencing mode can be achieved by delivery of Mn^{++} to the sequencing reaction, such that the final concentration of Mn^{++} is about 300 μM or more, about 400 μM or more, about 500 μM or more, about 700 μM or more, or about 1 mM or more.

[0139] In a preferred embodiment, the sequencing reactions take place in a structure that provides optical confinement, e.g., a nanohole or zero-mode waveguide. Further details regarding confinement strategies, substrates and systems for monitoring sequencing reactions can be found in co-pending published U.S. Patent Application No. 2007-0188750, and published International Patent Application No. WO 2007/095119, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[0140] The sequencing methods and systems of the present invention can employ a wide variety of template nucleic acids. In some cases, the template nucleic acid is a linear template. For a linear template, switching between a high accuracy and a high read length mode can result in a read wherein a relatively long stretch of template nucleic acid is sequenced in a given time, wherein generally, multiple regions along the length will be sequenced with high accuracy. The knowledge that the high accuracy regions are arranged in a particular sequential manner can be useful in putting together the sequence of the nucleic acid. In some cases, the template nucleic acid is a circular template. For a circular template, switching between high accuracy and long read length modes can result in the same region of the template nucleic acid sequenced by each of the modes. In a preferred embodiment, the template to be sequenced in more than one sequencing mode is a single-stranded nucleic acid loop. Double-stranded templates can reanneal, reducing primer annealing efficiency and impeding the polymerase-catalyzed extension of a sequencing reaction. In fact, loops can also be preferable to linear templates because a DNA polymerase can only copy a linear template, e.g., to which a primer has been annealed, once before it falls off the distal end of the template. In contrast, a strand-displacing polymerase can replicate a contiguous nucleic acid loop several times. The primer that is annealed to the loop is eventually displaced at its 5'-end upon completion of one revolution of the polymerase around the nucleic acid loop, and as

polymerization and displacement continue, a linear, single-stranded product comprising several copies of the nucleic acid sequence of the loop is generated. Accordingly, using nucleic acid loops in sequencing can provide an internal sequencing control.

[0141] The methods for preparing closed, single-stranded nucleic acid loops include providing a genomic DNA, a cDNA, or a DNA concatamer and generating double-stranded fragments that each comprise a first strand (e.g., an exonuclease sensitive strand) and a second strand (e.g., an exonuclease resistant strand). In a following step, the two strands in each fragment are separated, and the resulting single-stranded fragments are circularized to produce closed single-stranded nucleic acid loops, which can then be used as templates in a high-throughput sequencing system. Further details regarding the preparation of single stranded nucleic acid loops, and their use as templates in high-throughput sequencing systems, can be found in Patel et al. PCT Application Serial No. PCT/US2009/001930 METHODS AND COMPOSITIONS FOR NUCLEIC ACID SAMPLE PREPARATION, incorporated herein by reference in its entirety for all purposes.

[0142] Systems for analyzing the data generated during sequencing of a template nucleic acid using more than one sequencing mode are also a feature of the invention. Such systems will include a signal detector, e.g., in the case of a plurality of arrayed sequencing reactions, an array detector, e.g., an EMCCD. The detector, is then operatively coupled to a data storage and processing system. In a first sequencing mode, e.g., a mode where a high branching fraction is induced, the processing system is capable of interpreting multiple, e.g.,. redundant, or iterative signal pulses for each nucleotide incorporation event during a sequencing reaction to call bases with increased accuracy. In a subsequent sequencing mode, e.g., a high read length mode in which little or no branching occurs, the processing system is capable of interpreting single signal pulses as nucleotide incorporation events. Further details regarding base calling during sequencing by incorporation methods are found in Tomaney et al. PCT Application Serial No. PCT/US2008/065996 METHODS AND PROCESSES FOR CALLING BASES IN SEQUENCING BY INCORPORATION METHODS, incorporated herein by reference in its entirety for all purposes.

[0143] The multi-modal sequencing method of the present invention optionally includes recording signals or signatures, e.g., in a computer readable medium. The signals or signatures can be stored, e.g., as graphic or digital information. Any typical recording device such as a hard drive, memory card, memory stick, optical storage device or floppy drive can be used to record detected signals. Signals or signal signatures can also be deconvoluted or translated to provide, e.g., sequence information, e.g., in sequencing systems of the invention. Signal processing equipment can include, e.g., a computer having appropriate software for converting signals into sequence or assay parameter information.

[0144] Signal detection optics can be coupled to cameras, digital processing apparatus, or the like, to record and analyze signals detected in the various systems herein, . Systems can include a microscope, a CCD, a phototube, a photodiode, an LCD, a scintillation counter, film for recording signals, and the like. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image, e.g., using PC (Intel x86 or pentium chip-compatible DOS™, OS2™ WINDOWS™, WINDOWS NT™ or WINDOWS95™ based machines), MACINTOSH™, LINUX, or UNIX based (e.g., SUN™ work station) computers or digital appliances. Computers and digital appliances can include software for analyzing and perfecting signal interpretation. This can include standard application software such as spreadsheet or database software for storing signal information. However, systems of the invention can also include statistical analysis software to interpret signal or signature data, e.g., to translate the data into nucleic acid sequence information. For example, many vendors, such as Partek Incorporated (St. Peters, Mo.; www.partek.com) provide software for pattern recognition which can be applied to signal interpretation and analysis. Algorithms for sequencing systems that can be adapted to the invention are also described in Tomaney et al. PCT Application Serial No. PCT/US2008/065996 METHODS AND PROCESSES FOR CALLING BASES IN SEQUENCING BY INCORPORATION METHODS, incorporated herein by reference in its entirety for all purposes. Once signal information has been converted into sequence information, standard sequence aalysis software can be used to assemble overlapping sequence information. For example, sequence contigs can be assembled using available software such as DNA Baser (Heracle Software, Germany), or Artemis 11 (Sanger Institute) "Artemis and ACT:

Viewing, annotating and comparing sequences stored in a relational database" Carver et al. Bioinformatics 2008 PMID: 18845581 DOI: 10.1093/bioinformatics/btn529).

[0145] Relationships between datasets (e.g., high accuracy versus high readlength data) can similarly be analyzed, e.g., by pattern recognition software, Bayes classifiers, neural networks, Monte Carlo analysis, Principal Component Analysis (PCA), etc. Further information regarding genetic algorithms and neural networks that can be used to analyze signal or signature information can be found in David E. Goldberg (1989) Genetic Algorithms in Search, Optimization and Machine Learning Addison-Wesley Pub Co; ISBN: 0201157675; Timothy Masters (1993) Practical Neural Network Recipes in C++(Book & Disk edition) Academic Pr; ISBN: 0124790402; Kevin Gurney (1999) An Introduction to Neural Networks, UCL Press, 1 Gunpowder Square, London EC4A 3DE, UK; Christopher M. Bishop (1995) Neural Networks for Pattern Recognition Oxford Univ Press; ISBN: 0198538642; Brian D. Ripley, N. L. Hjort (Contributor) (1995) Pattern Recognition and Neural Networks Cambridge Univ Pr (Short); ISBN: 0521460867;Rubinstein, R. Y.; Kroese, D. P. (2007) Simulation and the Monte Carlo Method (2nd ed.). New York: John Wiley & Sons. ISBN 9780470177938; Tarantola, Albert (2005) Inverse Problem Theory Philadelphia: Society for Industrial and Applied Mathematics ISBN 0898715725; Steeb (2008) The Nonlinear Workbook: Chaos, Fractals, Neural Networks, Genetic Algorithms, Gene Expression Programming, Support Vector Machine, Wavelets, Hidden Markov Models, Fuzzy Logic with C++, Java and SymbolicC++ Programs: 4th edition. World Scientific Publishing. ISBN 981-281-852-9; Sergios Theodoridis, Konstantinos Koutroumbas, (2009) Pattern Recognition (4th edition), Elsevier, ISBN 978-1-59749-272-0, and in a variety of other currently available references. Computers/digital appliances also optionally include or are operably coupled to user viewable display systems (monitors, CRTs, printouts, etc.), printers to print data relating to signal information, peripherals such as magnetic or optical storage drives, user input devices (keyboards, microphones, pointing devices) and the like.

[0146] One example of a system that utilizes more than one mode of sequencing is illustrated in **Figure 7**. As shown, system 700 includes substrate 702 of the invention, e.g., upon which sequencing reactions are performed. Optical energy source 704 and additional optical energy source 706 deliver excitation light to the substrate 702, via an optical train. As

shown, the optical train includes a dichroic mirror 708 that is transmissive to excitation wavelength of the light from laser 704, while being reflective of light of the wavelength produced by laser 706, allowing both illumination beams to be in the same path. The excitation light is then directed at a second dichroic mirror 710, that reflects the excitation light through objective lens 712, and onto the substrate 702. Optical signals from the substrate 702 are then collected by objective lens 712, and are passed through dichroic 710. The fluorescent signals are then subjected to spatial separation using, e.g., a dispersive optical element, such as an optical grating or prism 714. The separated signals are then focused upon array detector 716, e.g., an EMCCD. The detector, is then operatively coupled to a data storage and processing system, such as computer 718 for processing and storage of the signal data and presentation of the data in a user desired format, e.g., on printer 720.

[0147] As shown in Figure 7, in one example of multi-modal sequencing, a first sequencing mode (Panel I) is produced in which a sequencing by incorporation reaction occurs under higher branch fraction conditions than a second sequencing mode. During the first sequencing mode, the processing system of the computer is capable of interpreting multiple, e.g.,. redundant, or iterative signal pulses or signatures for each nucleotide incorporation event during a sequencing reaction to call bases. Subsequently, a second sequencing mode (Panel II) is produced in which a sequencing by incorporation reaction occurs with no or very few branch fraction nonincorporation events, and accordingly, longer readlengths. During the second sequencing mode, the processing system of the computer interprets one or very few signal pulses or signatures as a nucleotide incorporation event during the sequencing reaction to call bases. The system can utilize signal pulse or signature data from each sequencing mode in order to determine the sequence of the template nucleic acid. As will be appreciated, a portion of the nucleic acid template can be sequenced once (in one of the two or more sequencing modes) or two or more times (two or more times in the same or multiple sequencing modes).

[0148] Although illustrated as an optical train that is transmissive of fluorescent signals, e.g., as provided by dichroic **710**, it will be appreciated that fluorescence reflective systems may also be employed. Further details regarding base calling during sequencing by incorporation methods are found in Tomaney et al. PCT Application Serial No. PCT/US2008/065996 METHODS AND PROCESSES FOR CALLING BASES IN

SEQUENCING BY INCORPORATION METHODS, incorporated herein by reference in its entirety for all purposes.

[0149] As will be appreciated, sequencing of a nucleic acid template in more than one sequencing mode can be combined with (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (3) iterative sampling of unincorporatable nucleotides; (4) two slow-step enzyme systems; (5) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (6) modified recombinant polymerases that exhibit increased nucleotide residence time as compared to the corresponding wild-type polymerases; and/or (7) any other combination of embodiments described herein.

III. DELAYED TRANSLOCATION

[0150] The present invention also provides polymerases that exhibit delayed translocation as compared to parental/wild type enzymes. Time necessarily lapses between the incorporation of one nucleotide and the incorporation of the next nucleotide due to the sequential (rather than simultaneous) nature of nucleotide incorporation. The duration of this time lapse is determined primarily by the rate at which a polymerase translocates along a template polynucleotide between incorporation events ("translocation" refers to the movement of a DNA polymerase along a template polynucleotide from an initial enzyme binding site to a subsequent enzyme binding site, where the enzyme binding sites correlate to nucleotide incorporation sites). Upon incorporation of a nucleotide, a polymerase is unable to accept another nucleotide until it has gone through the translocation process and moved into the next incorporation site.

[0151] When a polymerase translocates at a typical wild-type rate from one incorporation site to the next, a single molecule sequencing (SMS) system has less time to identify and distinguish when the polymerase has moved into the next incorporation site because the rapid translocation affords little separation between the signal pulses arising from incorporation events at one incorporation site and the pulses arising from those events at the next incorporation site. Accordingly, SMS under conditions of typical polymerase

translocation rates can have an increased potential for deletion errors relative to the correct template sequence (i.e., signals from incorporation events may not be detected or may be detected but not interpreted as incorporation events). These deletions would be artifacts that constitute errors in sequencing reads.

[0152] In accordance with aspects of the present invention, however, the modified recombinant polymerases provided by the present invention that exhibit a translocation delay of longer duration as compared to parental (or wild-type) polymerases are advantageous in the context of SMS because they provide greater temporal separation (i.e., resolution) between the signal pulse emitted while the polymerase resides at one incorporation site from the pulse emitted once the polymerase has entered the next incorporation site. This increased translocation delay permits the sequencing system to interpret signal pulses as correlating to separate and distinct nucleotide incorporation events, and hence greatly diminishes the potential for missed pulses or inaccurately characterized pulses relative to the correct template sequence. Further, a delay in translocation enhances the ability of a sequencing system to distinguish branching events that exhibit short inter-pulse widths from incorporation events that, under the conditions of delayed translocation, exhibit markedly broader inter-pulse widths.

[0153] In one aspect, the polymerases of the present invention are engineered to exhibit two sequential translocation kinetic steps that occur at a slower rate as compared to a parental polymerase, e.g., a wild-type parental polymerase. The two translocation kinetic steps may occur at the same rate, or the ratio of their rates may vary up to about 1:5 or 5:1 or more. The occurrence of exceedingly short inter-pulse widths, e.g., the time delay between signal pulses derived from sequential nucleotide incorporation events, is markedly reduced when two slow translocation kinetic steps are engineered into the polymerase. However, the present invention also provides polymerases that exhibit a translocation delay arising from a slow translocation kinetic step in conjunction with a slow kinetic step not related to translocation, e.g., a slow nucleotide and/or nucleotide analog binding step. Further details regarding polymerase systems with two slow kinetic steps can be found in Bjornson et al. PCT Application Serial Number PCT/US2009/002003 TWO SLOW-STEP POLYMERASE ENZYME SYSTEMS AND METHODS, incorporated herein by reference in its entirety for all purposes.

[0154] The delay in translocation exhibited by the polymerases of the invention can be more than 2.5x greater, e.g., more than 5x greater, more than 10x greater, more than 15x greater, more than 50x greater, more than 100x greater, more than 1000x greater, or more than 10,000x greater as compared to a parental polymerase (e.g., a wild type Φ29 polymerase) or more. Modified recombinant polymerases of the invention can allow the translocation step of a template-dependent polymerization reaction to be observable, e.g., wherein an otherwise unobservable translocation step (e.g., due to a high rate of translocation of an unmodified polymerase and limitations of the detection system) becomes observable as a result of modification of the polymerase.

[0155] The present invention also provides modified recombinant polymerases that include a heterologous polypeptide sequence fused at or near the c-terminus of the polymerase and/or amino acid substitutions or deletions, in order to delay translocation of the polymerase. As noted above, when a polymerase translocates at a typical wild-type rate from one incorporation site to the next, a single molecule sequencing (SMS) system can fail to distinguish when the polymerase has moved into the next incorporation site because the rapid translocation affords little separation between the signal pulses arising from branching and incorporation events at one incorporation site and the pulses arising from those events at the next incorporation site. Accordingly, SMS under conditions of typical polymerase translocation rates is susceptible to incorrect insertions relative to the correct template sequence (i.e., multiple signal pulses arising from branching and incorporation events at a single incorporation site may be interpreted as pulses arising from incorporation events at more than one incorporation site). These insertions are artifacts that constitute errors in sequencing reads.

[0156] The present invention provides modified or recombinant DNA polymerases that exhibit decreased translocation rates in order to provide greater temporal separation (i.e., resolution) between the signal pulses emitted while the polymerase resides at one incorporation site from the pulses emitted once the polymerase has entered the next incorporation site. This decreased translocation rate permits the sequencing system to interpret clusters of signal pulses as correlating to nucleotide incorporation sites, and hence greatly diminishes the occurrence of incorrect insertions relative to the correct template sequence.

[0157] In one embodiment, delayed translocation is accomplished by encoding a Φ29 DNA polymerase with a heterologous polypeptide sequence fused to the c-terminus of the DNA polymerase. For the purposes of this disclosure, "heterologous" refers to a polypeptide sequence that is not present in the wild-type parental polymerase. That these polymerases retain their functionality is a surprising aspect of the invention. The active site of the polymerase is located in the c-terminal portion of the protein, and previous attempts to modify the c-terminal portion have rendered the polymerase inactive. In one aspect, the heterologous polypeptide sequence can include between 6 and 10 positively charged amino acids, e.g., histidine. This stretch of positively charged amino acids can be encoded immediately downstream of the c-terminus of the polymerase. An example polymerase of this embodiment, comprising a Φ29 polymerase fused to 10 histidine residues at its c-terminus (SEQ ID No. 1), exhibits a translocation rate of 84 sec^{-1}, which constitutes a 2.3x delay in translocation as compared to an unmodified parental Φ29 polymerase under identical conditions. In another aspect, an amino acid linker sequence, e.g., a Ser3Gly linker (e.g., Gly-Gly-Gly-Ser-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Ser-Gly) is encoded between the stretch of positively charged amino acids and the C-terminus of the polymerase. An example polymerase of this embodiment, comprising a Φ29 polymerase fused to 10 histidine residues at its c-terminus, with a Ser3Gly linker between the polymerase and the histidine residues (SEQ ID No. 2), exhibits a translocation rate of 220 sec^{-1}, which constitutes a 2.4x delay in translocation as compared to an unmodified parental Φ29 polymerase under identical conditions. The affinity of positively charged amino acid residues at the c-terminus of the polymerase for the negatively charged phosphate backbone of DNA decreases the efficiency of polymerase translocation to the next nucleotide incorporation site, thereby causing a delay in translocation. Polymerases of the invention can also exhibit improved polymerase processivity, as the interaction between the polymerase and the phosphate backbone of DNA is enhanced by, e.g., the addition of positively-charged amino acid residues at the c-terminus.

[0158] In another embodiment, the polymerases that exhibit delayed translocation comprise amino acid substitutions and/or deletions that modulate interaction of the polymerases with the negatively-charged phosphate backbone of DNA. A number of specific examples of a modified polymerase, e.g. modified to delay polymerase translocation, are described herein. A

region of the polymerase responsible for interaction with the DNA template and primer is referred to herein as the binding cleft. Within the binding cleft are particular amino acid residues that interact with the phosphate backbone of DNA. The phosphate backbone constitutes a uniform interaction platform, e.g., nucleobase independent, and interactions with the phosphate backbone are altered in polymerases of the present invention, e.g., utilizing the negative charge of the backbone. For example, the interaction between a polymerase of the invention and the phosphate backbone of DNA can be enhanced by substituting neutral or negatively charged amino acids in the binding cleft with positively charged residues. However, mutations that delay polymerase translocation are not limited to this region of the polymerase. Relative to a wild-type Φ29 DNA polymerase, polymerase modifications of the present invention can include, e.g., any of the following mutations or combination of the following mutations: Asp570Lys; Asp570Ala; Asn313Lys; Asn313Ala; Gln303Lys; Gln303Ala; Gly532Ser; Met533delet; Cys530delet; Met533delet and Cys530delet; Gly532delet; Ala531Gly; Gly532Ser; Thr573Lys; Thr573Ala; Asn396Lys; Thr571Lys; Thr571Ala; Thr534Lys; Thr534Ala; Asp535Lys; Asp534Ala; Lys529Ala; and Lys529Asn. For the purposes of the present application, a mutation that includes a deletion at a particular residue position is presented by the amino acid abbreviation, followed by the residue position, followed by "del". Thus, for example, the Met533delet mutation presented above will be understood to mean a mutant polymerase in which the methionine at position 533 has been deleted.

[0159] The translocation delay exhibited by modified polymerases of the present invention, e.g., a modified Φ29 polymerase or a modified Φ 29-type polymerase, can be greater than the translocation delay exhibited by the parental polymerase or, e.g., about 2.5x greater, about 5x greater, about 10x greater, about 15x greater, about 50x greater or about 100x or more greater, as compared to a parental polymerase (e.g., a wild type Φ29 polymerase) under the standard conditions described above.

[0160] As will be appreciated, the above-identified modified or recombinant polymerases that display delayed translocation may optionally include additional modifications that confer other useful properties described herein, e.g., increased branching fraction, increased nucleotide analog residence time and/or increased processivity, etc. As will also be

appreciated, the above-identified modified or recombinant polymerases that display delayed translocation can be used in combination with any other embodiments described herein, including: (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (3) iterative sampling of unincorporatable nucleotides; (4) two slow-step enzyme systems; (5) detection of noncognate branching events; (6) modified recombinant polymerases that exhibit increased nucleotide residence time; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein.

IV. INCREASED RESIDENCE TIME

[0161] The present invention also provides modified recombinant polymerases that exhibit increased nucleotide or nucleotide analog residence time at an active site of the polymerase. During SMS, a number of signal parameters may be and generally are used for pulse identification. Two primary parameters are pulse intensity and pulse width, where pulse width relates to the duration of a signal pulse as detected by a signal detector of the sequencing system. Signal pulses that exhibit a short pulse width can arise from undesirable sources during SMS, e.g., transient pulses from labeled analogs in the reaction region that are not involved in an incorporation event. Incorporation events, on the other hand, are generally characterized by longer pulse durations stemming from increased residence time of the labeled nucleotide in the observation region by virtue of it being complexed with the polymerase.

[0162] Notwithstanding the differences in residence time duration upon incorporation, in some cases it would be desirable to further increase residence time for incorporated nucleotides, in order to further enhance the distinction between incorporation and transient signal events.

[0163] The modified recombinant polymerases of the present invention improve the accuracy of template-dependent polymerization reactions by increasing the residence times of nucleotides or nucleotide analogs at an active site of the polymerase, thereby producing signal pulses of increased width to further facilitate identification as signal pulses resulting from incorporation events.

[0164] As noted above, signal pulse intensity and width are primary parameters for signal pulse identification during template-dependent polymerization reactions, e.g., SMS utilizing dye-labeled nucleotide analogs. The signal pulse width is largely determined by the residence time of the nucleotide analog at the active site of the polymerase during the nucleotide incorporation step of the polymerase kinetic cycle. Insufficient residence times produce signal pulses of short duration and, accordingly, narrow pulse width that can prevent detection of genuine nucleotide incorporation events by a signal detector of the sequencing system. Undetected incorporation events arising from insufficient residence time and, accordingly, signal pulse width, constitute sequencing errors that diminish the reliability of SMS results.

[0165] The present invention provides modified or recombinant polymerases that improve sequence read accuracy by increasing the residence time of a nucleotide analog at the polymerase active site. During the polymerase kinetic cycle, interaction between a DNA polymerase and a nucleotide or nucleotide analog induces a conformational change in the polymerase, referred to herein as polymerase isomerization. During polymerase isomerization, the nucleotide or nucleotide analog is effectively immobilized at the active site of the polymerase. The duration of the isomerization step, therefore, affects the residence time of the nucleotide analog at the polymerase active site, and, accordingly, affects the signal pulse width as detected by a signal detector of a sequencing system. A slow isomerization step, therefore, can increase signal pulse width and enhance the accuracy of applications such as SMS. Further details regarding DNA polymerase enzymes and/or nucleotides or nucleotide analogs that provide altered residence times for enhanced nucleic acid sequence analysis and determination can be found in Rank et al. U.S. Application Serial No. 11/977,160 POLYMERASE ENZYMES AND REAGENTS FOR ENHANCED NUCELIC ACID SEQUENCING, incorporated herein by reference in its entirety for all purposes.

[0166] Completion of the nucleotide or nucleotide analog incorporation step of the DNA polymerase kinetic cycle requires dissociation of the polymerase from the incorporating nucleotide. Until dissociation occurs, the nucleotide, e.g., a dye-labeled nucleotide analog, is effectively immobilized at the active site of the polymerase. Slowing the dissociation step, therefore, results in increased nucleotide residence time at the active site of the polymerase and

increases the duration of a signal pulse emitted by a dye-labeled nucleotide analog as it is incorporated into the copy nucleic acid.

[0167] Polymerases of the present invention increase the residence time of nucleotides or nucleotide analogs by slowing the isomerization and/or dissociation steps of the polymerase kinetic cycle. Accordingly, these polymerases facilitate signal pulse detection by increasing signal pulse width. For example, a polymerase provided by the present invention comprises the amino acid substitution T368P and exhibits an average nucleotide residence time that is about 2x greater as compared to the residence time of a wild-type parental polymerase under identical conditions.

[0168] The nucleotide or nucleotide analog residence time exhibited by modified polymerases of the present invention, e.g., a modified Φ29 polymerase or a modified Φ 29-type polymerase, can be greater than the residence time exhibited by the parental polymerase or, e.g., about 1.5x greater, about 2.0x greater or about 3.0x or more greater, under the standard conditions described above.

[0169] As will be appreciated, the above-identified modified or recombinant polymerases that display increased nucleotide or nucleotide analog residence time may optionally include additional modifications that confer other useful properties described herein, e.g., increased branching fraction, delayed translocation and/or increased processivity. As will also be appreciated, the above-identified modified or recombinant polymerases that display increased nucleotide or nucleotide analog residence time can be used in combination with any other embodiments described herein, including: (1) reaction conditions that increase the frequency of branching fraction nonincorporation events; (2) modified recombinant polymerases that exhibit increased branching fractions as compared to the corresponding wild-type polymerases; (3) iterative sampling of unincorporatable nucleotides; (4) two slow-step enzyme systems; (5) detection of noncognate branching events; (6) modified recombinant polymerases that exhibit altered translocation properties as compared to the corresponding wild-type polymerases; (7) sequencing of nucleic acid templates using more than one mode; and/or (8) any other combination of embodiments described herein.

[0170] The properties of increased branching rates, delayed translocation and increased nucleotide or nucleotide analog residence time are particularly useful in the context of an incorporation of labeled nucleotides by the polymerase, e.g., as detected during sequencing by incorporation methods (including, e.g., SMS methods). For example, the invention provides, e.g., compositions that include one or more engineered or modified polymerase enzymes optionally with one or more template DNAs, and/or labeled or otherwise modified nucleotides or nucleotide analogs, where the composition exhibits increased branching rates and/or delayed polymerase translocation during template dependent polymerase-mediated nucleic acid synthesis. Methods, including SMS using these compositions, are also provided, as are general methods of making polymerases having the properties noted herein.

[0171] Accordingly, among other aspects, the present invention provides new polymerases that incorporate nucleotide analogs, such as phosphate labeled analogs, into a growing template copy during DNA amplification. These polymerases are modified such that they have increased branching rates and/or delayed translocation and/or increased residence time when incorporating the relevant analogs, and optionally have improved DNA-polymerase processivity as compared to corresponding wild-type parental polymerases (e.g., polymerases from which modified recombinant polymerases of the invention were derived, e.g., by mutation).

[0172] These new polymerases and reaction conditions are particularly well suited to DNA amplification and/or sequencing applications, particularly sequencing protocols that include detection in real time of the incorporation of labeled analogs into DNA amplicons, because the increased branching rate, delayed translocation and/or nucleotide residence time facilitates the correct determination of which labeled nucleotide is incorporated at a site during a template dependent polymerization reaction.

FURTHER DETAILS REGARDING SYSTEMS FOR SEQUENCING BY INCORPORATION

[0173] One example of a system for use in the present invention is illustrated in **Figure 1**. As shown, system 100 includes substrate 102 of the invention, e.g., upon which sequencing reactions are performed. Optical energy source 104 and additional optical energy source 106

deliver excitation light to the substrate 102, via an optical train. As shown, the optical train includes a dichroic mirror 108 that is transmissive to excitation wavelength of the light from laser 104, while being reflective of light of the wavelength produced by laser 106, allowing both illumination beams to be in the same path. The excitation light is then directed at a second dichroic mirror 110, that reflects the excitation light through objective lens 112, and onto the substrate 102. Optical signals from the substrate 102 are then collected by objective lens 112, and are passed through dichroic 110. The fluorescent signals are then subjected to spatial separation using, e.g., a dispersive optical element, such as an optical grating or prism 114. The separated signals are then focused upon array detector 116, e.g., an EMCCD. The detector, is then operatively coupled to a data storage and processing system, such as computer 118 for processing and storage of the signal data and presentation of the data in a user desired format, e.g., on printer 120. The processing system of the computer is capable of interpreting multiple, e.g.,. redundant, or iterative signal pulses for each nucleotide incorporation event during a sequencing reaction to call bases with increased accuracy. Although illustrated as an optical train that is transmissive of fluorescent signals, e.g., as provided by dichroic 110, it will be appreciated that fluorescence reflective systems may also be employed. Further details regarding base calling during sequencing by incorporation methods are found in Tomaney et al. PCT Application Serial No. PCT/US2008/065996 METHODS AND PROCESSES FOR CALLING BASES IN SEQUENCING BY INCORPORATION METHODS, incorporated herein by reference in its entirety for all purposes.

[0174] As will be appreciated, a number of other components may be included in the systems described herein, including optical filters for filtering background illumination or bleed-through illumination from the light sources, from the actual optical signals. Additionally, alternate optical trains may employ cascaded spectral filters in separating different spectral signal components. A monitor of the computer can display optical signal pulse patterns 122 generated under the conditions provided by the invention, e.g., increased branching, delayed polymerase translocation or increased nucleotide analog residence time. A variety of other optical configurations may additionally be employed in conjunction with the compositions of the invention.

[0175] In the context of nucleic acid sequencing methods, it will be appreciated that the signal sources each represent sequencing reactions, and particularly, polymerase-mediated, template-dependent primer extension reactions, where in preferred aspects, each base incorporation event results in a prolonged illumination (or localization) of one of four differentially labeled nucleotides being incorporated, so as to yield a recognizable pulse or signature that carries a distinguishable spectral profile or color.

[0176] In the context of the present invention, a signal event is first classified as to whether it constitutes a significant signal pulse or signature based upon whether such signal event meets any of a number of different criteria. Once identified or classified as a significant pulse or signature, the signal pulse or signature may be further assessed to determine whether the signal pulse or signature constitutes an incorporation event and may be called as a particular incorporated base. As will be appreciated, the basis for calling a particular signal event as a significant pulse or signature, and ultimately as an incorporation event, will be subject to a certain amount of error, based upon a variety of parameters as generally set forth herein. The reaction conditions and modified recombinant polymerases of the present invention diminish the error associated with base calling.

[0177] Once a particular signal is identified as a significant pulse or signature and is assigned a particular spectrum, e.g. color, the spectrally assigned pulse may be further assessed to determine whether the pulse or signature can be called an incorporation event and, as a result, call the base incorporated in the nascent strand, or its complement in the template sequence. Calling of bases from color assigned pulse or signature data will typically employ tests that identify the confidence level with which a base is called. Typically, such tests will take into account the data environment in which a signal was received, including a number of the same data parameters used in identifying significant pulses, etc. For example, such tests may include considerations of background signal levels, adjacent pulse signal parameters (spacing, intensity, duration, etc.), spectral image resolution, and a variety of other parameters. Such data may be used to assign a score to a given base call for a color assigned signal pulse or signature, where such scores are correlative of a probability that the base called is incorrect, e.g., 1 in 100 (99% accurate), 1 in 1000 (99.9% accurate), 1 in 10,000 (99.99% accurate), 1 in 100,000 (99.999% accurate), or even greater.

[0178] Once a base is called with sufficient accuracy, subsequent bases called in the same sequencing run, and in the same primer extension reaction, may then be appended to each previously called base to provide a sequence of bases in the overall sequence of the template or nascent strand. Iterative processing and further data processing, as described in greater detail below, can be used to fill in any blanks, correct any erroneously called bases, or the like for a given sequence.

POLYMERASES AND NUCLEOTIDE ANALOGS

[0179] Various polymerases may be used in the methods, compositions and systems described herein, including DNA polymerases, RNA polymerases, reverse transcriptases, and mutant or altered forms of any of the foregoing.

[0180] DNA polymerases that can be modified to increase the average branching fraction, decrease the translocation rate or increase nucleotide residence time are generally available. DNA polymerases are sometimes classified into six main groups based upon various phylogenetic relationships, e.g., with *E. coli* Pol I (class A), *E. coli* Pol II (class B), *E. coli* Pol III (class C), Euryarchaeotic Pol II (class D), human Pol beta (class X), and *E. coli* UmuC/DinB and eukaryotic RAD30/xeroderma pigmentosum variant (class Y). For a review of recent nomenclature, see, e.g., Burgers et al. (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" J Biol Chem. 276(47):43487-90. For a review of polymerases, see, e.g., Hübscher et al. (2002) "Eukaryotic DNA Polymerases" Annual Review of Biochemistry Vol. 71: 133-163; Alba (2001) "Protein Family Review: Replicative DNA Polymerases" Genome Biology 2(1):reviews 3002.1-3002.4; and Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" J Biol Chem 274:17395-17398. The basic mechanisms of action for many polymerases have been determined. The sequences of literally hundreds of polymerases are publicly available, and the crystal structures for many of these have been determined, or can be inferred based upon similarity to solved crystal structures for homologous polymerases. For example, the crystal structure of Φ29, a preferred type of parental enzyme to be modified according to the invention, is available.

[0181] DNA polymerases and their properties are described in detail in, among other places, DNA Replication 2nd Edition, Komberg and Baker, W.H. Freeman, New York, N.Y.

(1991). Known conventional DNA polymerases useful in the invention include, but are not limited to, Pyrococcus furiosus (Pfu) DNA polymerase (Lundberg et al., 1991, Gene, 108: 1, Stratagene), Pyrococcus woesei (Pwo) DNA polymerase (Hinnisdaels et al., 1996, Biotechniques, 20:186-8, Boehringer Mannheim), Thermus thermophilus (Tth) DNA polymerase (Myers and Gelfand 1991, Biochemistry 30:7661), Bacillus stearothermophilus DNA polymerase (Stenesh and McGowan, 1977, Biochim Biophys Acta 475:32), Thermococcus litoralis (Tli) DNA polymerase (also referred to as Vent.TM. DNA polymerase, Cariello et al., 1991, Polynucleotides Res, 19: 4193, New England Biolabs), 9.degree. Nm.TM. DNA polymerase (New England Biolabs), Stoffel fragment, ThermoSequenase.RTM. (Amersham Pharmacia Biotech UK), Therminator.TM. (New England Biolabs), Thermotoga maritima (Tma) DNA polymerase (Diaz and Sabino, 1998 Braz J Med. Res, 31:1239), Thermus aquaticus (Taq) DNA polymerase (Chien et al., 1976, J. Bacteoriol, 127: 1550), DNA polymerase, Pyrococcus kodakaraensis KOD DNA polymerase (Takagi et al., 1997, Appl. Environ. Microbiol. 63:4504), JDF-3 DNA polymerase (from thermococcus sp. JDF-3, Patent application WO 0132887), Pyrococcus GB-D (PGB-D) DNA polymerase (also referred as Deep Vent.TM. DNA polymerase, Juncosa-Ginesta et al., 1994, Biotechniques, 16:820, New England Biolabs), UlTma DNA polymerase (from thermophile Thermotoga maritima; Diaz and Sabino, 1998 Braz J. Med. Res, 31:1239; PE Applied Biosystems), Tgo DNA polymerase (from thermococcus gorgonarius, Roche Molecular Biochemicals), E. coli DNA polymerase I (Lecomte and Doubleday, 1983, Polynucleotides Res. 11:7505), T7 DNA polymerase (Nordstrom et al., 1981, J Biol. Chem. 256:3112), and archaeal DP11/DP2 DNA polymerase II (Cann et al., 1998, Proc Natl Acad. Sci. USA 95:14250-5).

[0182] Reverse transcriptases useful in the invention include, but are not limited to, reverse transcriptases from HIV, HTLV-1, HTLV-II, FeLV, FIV, SIV, AMV, MMTV, MoMuLV and other retroviruses (see Levin, Cell 88:5-8 (1997); Verma, Biochim Biophys Acta. 473:1-38 (1977); Wu et al., CRC Crit Rev Biochem. 3:289-347 (1975)).

[0183] In preferred embodiments, the polymerases employed during the sequencing processes, and optionally during pre-sequencing synthesis, will typically possess strand-displacement activity to displace any primers downstream of the primer at which the strand synthesis is initiated. A preferred rolling circle polymerase exhibits strand-displacement

activity, and as such, a single circular template can be sequenced repeatedly to produce a sequence read comprising multiple copies of the complement of the template strand by displacing the nascent strand ahead of the translocating polymerase. Non-limiting examples of rolling circle polymerases suitable for the present invention include but are not limited to T5 DNA polymerase, T4 DNA polymerase holoenzyme, phage M2 DNA polymerase, phage PRD1 DNA polymerase, Klenow fragment of DNA polymerase, and certain polymerases that are modified or unmodified and chosen or derived from the phages Φ29, PRD1, Cp-1, Cp-5, Cp-7, Φ15, Φ1, Φ21, Φ25, BS 32 L17, PZE, PZA, Nf, M2Y (or M2), PR4, PR5, PR722, B103, SF5, GA-1, and related members of the *Podoviridae* family. In certain preferred embodiments, the polymerase is a modified Phi29 DNA polymerase, e.g., as described in U.S. Patent Publication No. 20080108082, incorporated herein by reference in its entirety for all purposes. Similarly, polymerases having enhanced activity for labeled nucleotides are also desirable. Examples of polymerase enzymes for use in various aspects of the invention include, e.g., those described in U.S. Patent Application Nos. 11/645,125, filed December 21, 2006; 11/645,135, filed December 21, 2006; 12/384,112, filed March 30, 2009; 61/094,843, filed September 5, 2008; and 61/072,645, filed March 31, 2008; as well as U.S. Patent Publication No. 20070196846 (the full disclosures of which are incorporated herein by reference in their entireties for all purposes), such as the E375Y/K512Y/T368F mutant polymerase described in the foregoing.

[0184] In addition to wild-type polymerases, chimeric polymerases made from a mosaic of different sources can be used. For example, Φ29 polymerases made taking sequences from more than one parental polymerase into account can be used as a starting point for mutating the polymerases of the invention. This can done using consideration of similarity regions between the polymerases to define consensus sequences that are used in the chimera, or can be done using gene shuffling technologies in which multiple Φ29-related polymerases are randomly or semi-randomly shuffled via available gene shuffling techniques (e.g., via "family gene shuffling"; see Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288–291; Clackson et al. (1991) "Making antibody fragments using phage display libraries" Nature 352:624-628; Gibbs et al. (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of

recombination with family shuffling" Gene 271:13-20; and Hiraga and Arnold (2003) "General method for sequence-independent site-directed chimeragenesis: J. Mol. Biol. 330:287-296). In these methods, the recombination points can be predetermined such that the gene fragments assemble in the correct order. However, the combinations, e.g., chimeras, can be formed at random. For example, using methods described in Clarkson, et al., 5 gene chimeras, e.g., comprising segments of a Phi29 polymerase, a PZA polymerase, a M2 polymerase, a B103 polymerase, and a GA-1 polymerase, with improved branching fractions can be generated.

[0185] Available DNA polymerase enzymes have also been modified in any of a variety of ways, e.g., to simplify production by making protease digested enzyme fragments such as the Klenow fragment recombinant, etc. As noted, polymerases have also been modified to confer improvements in specificity, processivity, and improved retention time of labeled nucleotides in polymerase-DNA-nucleotide complexes (e.g., WO 2007/076057 POLYMERASES FOR NUCLEOTIDE ANALOG INCORPORATION by Hanzel et al., and PCT/US2007/022459 POLYMERASE ENZYMES AND REAGENTS FOR ENHANCED NUCLEIC ACID SEQUENCING by Rank et al.) and to improve surface-immobilized enzyme activities (e.g., WO 2007/075987 ACTIVE SURFACE COUPLED POLYMERASES by Hanzel et al., and WO 2007/076057 PROTEIN ENGINEERING STRATEGIES TO OPTIMIZE ACTIVITY OF SURFACE ATTACHED PROTEINS by Hanzel et al.). Any of these available polymerases can be modified in accordance with the invention to increase the average branching fraction, delay translocation and/or increase nucleotide or nucleotide analog residence time.

[0186] The polymerase mutations and mutational strategies noted herein can be combined with available mutations and mutational strategies to confer additional improvements in, e.g., nucleotide analog specificity, enzyme processivity and the like. For example, the mutations and mutational strategies herein can be combined with those taught in, e.g., WO 2007/076057 POLYMERASES FOR NUCLEOTIDE ANALOG INCORPORATION by Hanzel et al., and PCT/US2007/022459 POLYMERASE ENZYMES AND REAGENTS FOR ENHANCED NUCLEIC ACID SEQUENCING by Rank et al. This combination of mutations/ mutational strategies can be used to impart several simultaneous improvements to a polymerase (decreased branch fraction formation, improved specificity, improved processivity, improved

retention time, improved stability of the closed complex, etc.). In addition, polymerases can be further modified for application-specific reasons, such as to improve activity of the enzyme when bound to a surface, as taught, e.g., in WO 2007/075987 ACTIVE SURFACE COUPLED POLYMERASES by Hanzel et al., and WO 2007/076057 PROTEIN ENGINEERING STRATEGIES TO OPTIMIZE ACTIVITY OF SURFACE ATTACHED PROTEINS by Hanzel et al., or to include purification or handling tags as is taught in the cited references and as is common in the art.

[0187] Specific mutations noted herein can be used alone or in combination with each other and/or with available mutations as described in the references noted above, or can be used in polymerases that lack such previously described mutations.

[0188] Many such polymerases that are suitable for modification are available, e.g., for use in sequencing, labeling and amplification technologies. For example, Human DNA Polymerase Beta is available from R&D systems. DNA polymerase I is available from Epicenter, GE Health Care, Invitrogen, New England Biolabs, Promega, Roche Applied Science, Sigma Aldrich and many others. The Klenow fragment of DNA Polymerase I is available in both recombinant and protease digested versions, from, e.g., Ambion, Chimerx, eEnzyme LLC, GE Health Care, Invitrogen, New England Biolabs, Promega, Roche Applied Science, Sigma Aldrich and many others. Φ29 DNA polymerase is available from e.g., Epicentre. Poly A polymerase, reverse transcriptase, Sequenase, SP6 DNA polymerase, T4 DNA polymerase, T7 DNA polymerase, and a variety of thermostable DNA polymerases (Taq, hot start, titanium Taq, etc.) are available from a variety of these and other sources. Recent commercial DNA polymerases include Phusion™ High-Fidelity DNA Polymerase, available from New England Biolabs; GoTaq® Flexi DNA Polymerase, available from Promega; RepliPHI™ Φ29 DNA Polymerase, available from Epicentre Biotechnologies; PfuUltra™ Hotstart DNA Polymerase, available from Stratagene; KOD HiFi DNA Polymerase, available from Novagen; and many others. Biocompare(dot)com provides comparisons of many different commercially available polymerases.

[0189] DNA polymerases that are preferred substrates for mutation to increase branch rates and/or decrease translocation rate include Taq polymerases, *E. coli* DNA Polymerase 1,

Klenow fragment, reverse transcriptases, Φ29 related polymerases including wild type Φ29 polymerase and derivatives of such polymerases such as exonuclease altered forms, T7 DNA polymerase, T5 DNA polymerase, an RB69 polymerase, etc.

[0190] In one aspect, the modified recombinant polymerases that exhibit increased branching fractions, delayed translocation and/or increased nucleotide or nucleotide analog residence time are Φ29-type DNA polymerases. For example, the modified recombinant DNA polymerases can be homologous to a wild-type Φ29 DNA polymerase, e.g., as described in U.S. Patent Nos. 5,001,050, 5,198,543, or 5,576,204. Alternately, the modified recombinant DNA polymerase can be homologous to other Φ29-type DNA polymerases, such as B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, Φ21, or the like. For nomenclature, see also, Meijer et al. (2001) "Φ29 Family of Phages" Microbiology and Molecular Biology Reviews, 65(2):261-287.

[0191] As discussed, various polymerases of the invention can incorporate one or more nucleotide analogs into a growing oligonucleotide chain. Upon incorporation, the analog can leave a residue that is the same as or different than a natural nucleotide in the growing oligonucleotide (the polymerase can incorporate any non-standard moiety of the analog, or can cleave it off during incorporation into the oligonucleotide). A "nucleotide analog" herein is a compound, that, in a particular application, functions in a manner similar or analogous to a naturally occurring nucleoside triphosphate (a "nucleotide"), and does not otherwise denote any particular structure. A nucleotide analog is an analog other than a standard naturally occurring nucleotide, i.e., other than A, G, C, T, or U, though upon incorporation into the oligonucleotide, the resulting residue in the oligonucleotide can be the same as (or different from) an A, G, C, T, or U residue.

[0192] Nucleotide analogs can also be modified to achieve any of the improved properties desired. For example, various linkers or other substituents can be incorporated into analogs that have the effect of altering the branching rate, residence time or improving processivity. Modifications to the analogs can include extending the phosphate chains, e.g., to include a hexa- or heptaphosphate group, and/or adding chemical linkers to extend the distance between the nucleotide base and the dye molecule, e.g., fluorescent dye molecule. As

described in detail herein, modifications to the analogs can include altering the analog such that the analog is unincorporatable. For example, nucleotide analogs of the invention can possess unhydrolyzable groups within the phosphate chain, such that the phosphoester linkage between the analog and the primer strand cannot be formed.

[0193] Many nucleotide analogs are available and can be incorporated by the polymerases of the invention. These include analog structures with core similarity to naturally occurring nucleotides, such as those that comprise one or more substituents on a phosphate, sugar or base moiety of the nucleoside or nucleotide relative to a naturally occurring nucleoside or nucleotide. In one embodiment, the nucleotide analog includes three phosphate containing groups; for example, the analog can be a labeled nucleoside triphosphate analog and/or an α-thiophosphate nucleotide analog having three phosphate groups. In one embodiment, a nucleotide analog can include one or more extra phosphate containing groups, relative to a nucleoside triphosphate. For example, a variety of nucleotide analogs that comprise, e.g., from 4-6 or more phosphates are described in detail in U.S. Patent Application No. 11/241,809, filed September 29, 2005, and incorporated herein by reference in its entirety for all purposes. Other exemplary useful analogs, including tetraphosphate and pentaphosphate analogs, are described in U.S. Patent 7,041,812, incorporated herein by reference in its entirety for all purposes.

[0194] Nucleotide analogs of the present invention may comprise any of a variety of detectable labels. Detectable labels generally denote a chemical moiety that provides a basis for detection of the analog compound separate and apart from the same compound lacking such a labeling group. Examples of labels include, e.g., optical labels, e.g., labels that impart a detectable optical property to the analog, electrochemical labels, e.g., labels that impart a detectable electrical or electrochemical property to the analog, and physical labels, e.g., labels that impart a different physical or spatial property to the analog, e.g., a mass tag or molecular volume tag. In some cases individual labels or combinations may be used that impart more than one of the aforementioned properties to the analogs of the invention.

[0195] Optionally, the labeling groups incorporated into the analogs comprise optically detectable moieties, such as luminescent, chemiluminescent, fluorescent, fluorogenic, chromophoric and/or chromogenic moieties, with fluorescent and/or fluorogenic labels being

preferred. A variety of different label moieties are readily employed in nucleotide analogs. Such groups include fluorescein labels, rhodamine labels, cyanine labels (i.e., Cy3, Cy5, and the like, generally available from the Amersham Biosciences division of GE Healthcare), the Alexa family of fluorescent dyes and other fluorescent and fluorogenic dyes available from Molecular Probes/Invitrogen, Inc. and described in 'The Handbook — A Guide to Fluorescent Probes and Labeling Technologies, Tenth Edition' (2005) (available from Invitrogen, Inc./Molecular Probes). A variety of other fluorescent and fluorogenic labels for use with nucleoside polyphosphates, and which would be applicable to the nucleotide analogs incorporated by the polymerases of the present invention, are described in, e.g., U.S. Patent Application Publication No. 2003/0124576, previously incorporated herein by reference in its entirety for all purposes.

[0196] Additional details regarding analogs and methods of making such analogs can be found in U.S. Patent Application No. 11/241,809, filed September 29, 2005, and incorporated herein by reference in its entirety for all purposes.

[0197] Thus, in one illustrative example, the analog can be a phosphate analog (e.g., an analog that has more than the typical number of phosphates found in nucleoside triphosphates) that include, e.g., an Alexa dye label. For example, an Alexa488 dye can be labeled on a delta phosphate of a tetraphosphate analog (denoted, e.g., A488dC4P or A488dA4P, for the Alexa488 labeled tetraphosphate analogs of C and A, respectively), or an Alexa568 or Alexa633 dye can be used (e.g., A568dC4P and A633dC4P, respectively, for labeled tetraphosphate analogs of C or A568dT6P for a labeled tetraphosphate analog of T), or an Alexa546 dye can be used (e.g., A546dG4P), or an Alexa594 dye can be used (e.g., A594dT4P). An Alexa555 dye (A555dC6P), an Alexa 647 dye (A647d6GP) and/or an Alexa660 dye (A660dA6P) can be used in, e.g., single molecule sequencing. Similarly, to facilitate color separation, a pair of fluorophores exhibiting FRET (fluorescence resonance energy transfer) can be labeled on a delta phosphate of a tetraphosphate analog (denoted, e.g., FAM-amb-A532dG4P or FAM-amb-A594dT4P).

[0198] Polymerases of the invention that exhibit branching phenotypes will display a branching fraction that is particular to the nucleotide analog included in the polymerization

reaction. For example, a polymerase of the invention may exhibit different branching fractions for A488dC4P than A488dC6P due, e.g., to the size differences between 4P and 6P analogs. It will be appreciated that polymerases of the present invention can be modified such that it exhibits the desired branching fraction for a particular nucleotide analog, e.g., a dye-labeled nucleotide analog with a particular number of phosphate groups.

APPLICATIONS FOR ENHANCED NUCLEIC ACID SEQUENCING

[0199] Polymerases of the invention, e.g., modified recombinant polymerases, are used in combination with nucleotides and/or nucleotide analogs, and nucleic acid templates (DNA or RNA) to copy the template nucleic acid. That is, a mixture of the polymerase, nucleotides/analogs, and optionally and other appropriate reagents, the template and a replication initiating moiety (e.g., primer) is reacted such that the polymerase synthesizes nucleic acid (e.g., extends the primer) in a template-dependent manner. The replication initiating moiety can be a standard oligonucleotide primer, or, alternatively, a component of the template, e.g., the template can be a self-priming single stranded DNA, a nicked double stranded DNA, or the like. Similarly, a terminal protein can serve as a initiating moiety. At least one nucleotide analog can be incorporated into the DNA. The template DNA can be a linear or circular DNA, and in certain applications, is desirably a circular template (e.g., for rolling circle replication or for sequencing of circular templates). Optionally, the composition can be present in an automated DNA replication and/or sequencing system.

[0200] Incorporation of labeled nucleotide analogs by the polymerases of the invention is particularly useful in a variety of different nucleic acid analyses, including real-time monitoring of DNA polymerization. The label can itself be incorporated, or more preferably, can be released during incorporation of the analog. For example, analog incorporation can be monitored in real-time by monitoring label release during incorporation of the analog by the polymerase. The portion of the analog that is incorporated can be the same as a natural nucleotide, or can include features of the analog that differ from a natural nucleotide.

[0201] In general, label incorporation or release can be used to indicate the presence and composition of a growing nucleic acid strand, e.g., providing evidence of template replication/amplification and/or sequence of the template. Signaling from the incorporation can

be the result of detecting labeling groups that are liberated from the incorporated analog, e.g., in a solid phase assay, or can arise upon the incorporation reaction. For example, in the case of FRET labels where a bound label is quenched and a free label is not, release of a label group from the incorporated analog can give rise to a fluorescent signal. Alternatively, the enzyme may be labeled with one member of a FRET pair proximal to the active site, and incorporation of an analog bearing the other member will allow energy transfer upon incorporation. The use of enzyme bound FRET components in nucleic acid sequencing applications is described, e.g., in U.S. Patent Application Publication No. 2003/0044781, incorporated herein by reference.

[0202] In one example reaction of interest, a polymerase reaction can be isolated within an extremely small observation volume that effectively results in observation of individual polymerase molecules. In a preferred aspect, such small observation volumes are provided by immobilizing the polymerase enzyme within a structural confinement, such as a nanohole (an aperture of less than 1 μM diameter through which a synthesis complex can be illuminated by optical energy) (See, e.g., co-pending Published U.S. Patent Application No. 2007-0188750, and published International Patent Application No. WO 2007/095119, the full disclosures of which are incorporated herein by reference in their entirety for all purposes) or nanoholes that additionally provide optical confinement, such as a zero-mode waveguide (ZMW). For a description of ZMWs and their application in single molecule analyses, and particularly nucleic acid sequencing, see, e.g., U.S. Patent Application Publication No. 2003/0044781, and U.S. Patent No. 6,917,726, each of which is incorporated herein by reference in its entirety for all purposes. See also Levene et al. (2003) "Zero-mode waveguides for single-molecule analysis at high concentrations" Science 299:682-686 and U.S. Patent Nos. 7,056,676, 7,056,661, 7,052,847, and 7,033,764, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[0203] In general, a polymerase enzyme is complexed with the template strand in the presence of one or more nucleotides and/or one or more nucleotide analogs. For example, in certain embodiments, labeled analogs are present representing analogous compounds to each of the four natural nucleotides, A, T, G and C, e.g., in separate polymerase reactions, as in classical Sanger sequencing, or multiplexed together, e.g., in a single reaction, as in multiplexed sequencing approaches. When a particular base in the template strand is

encountered by the polymerase during the polymerization reaction, it complexes with an available analog that is complementary to such nucleotide, and incorporates that analog into the nascent and growing nucleic acid strand. In one aspect, incorporation can result in a label being released, e.g., in polyphosphate analogs, cleaving between the α and β phosphorus atoms in the analog, and consequently releasing the labeling group (or a portion thereof). The incorporation event is detected, either by virtue of a longer presence of the analog and, thus, the label, in the complex, or by virtue of release of the label group into the surrounding medium. Where different labeling groups are used for each of the types of analogs, e.g., A, T, G or C, identification of a label of an incorporated analog allows identification of that analog and consequently, determination of the complementary nucleotide in the template strand being processed at that time. Sequential reaction and monitoring permits a real-time monitoring of the polymerization reaction and determination of the sequence of the template nucleic acid. As noted above, in particularly preferred aspects, the polymerase enzyme/template complex is provided immobilized within an optical confinement that permits observation of an individual complex, e.g., a nanohole or zero mode waveguide.

[0204] In addition to their use in sequencing, the polymerases and/or reaction conditions of the invention are also useful in a variety of other genotyping analyses, e.g., SNP genotyping using single base extension methods, real time monitoring of amplification, e.g., RT-PCR methods, and the like. Further details regarding sequencing and nucleic acid amplification can be found, e.g., in Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2000; Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2006); and PCR Protocols: A Guide to Methods and Applications (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990).

MOLECULAR MODELING-BASED MODIFICATION OF POLYMERASES TO INCREASE BRANCHING FRACTION, DELAY TRANSLOCATION OR INCREASE NUCLEOTIDE ANALOG RETENTION TIME

Structure-based design of recombinant polymerases

[0205] Structural data for a polymerase can be used to conveniently identify amino acid residues as candidates for mutagenesis to create recombinant polymerases, e.g., having modified active site regions that increase the branching fractions. For example, analysis of the three dimensional structure of a polymerase such as Φ29 can identify residues that are particularly relevant to branching, translocation and/or nucleotide residence time properties of the polymerase.

[0206] The three-dimensional structures of a large number of DNA polymerases have been determined by x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, including the structures of polymerases with bound templates, nucleotides, and/or nucleotide analogs. Many such structures are freely available for download from the Protein Data Bank, at (www(dot)rcsb(dot)org/pdb. Structures, along with domain and homology information, are also freely available for search and download from the National Center for Biotechnology Information's Molecular Modeling DataBase, at www(dot)ncbi(dot)nlm(dot)nih(dot)gov/Structure/MMDB/mmdb(dot)shtml. The structures of additional polymerases can be modeled, for example, based on homology of the polymerases with polymerases whose structures have already been determined. Alternatively, the structure of a given polymerase, optionally complexed with a template and/or nucleotide analog, or the like, can be determined.

[0207] Techniques for crystal structure determination are well known. See, for example, McPherson (1999) Crystallization of Biological Macromolecules Cold Spring Harbor Laboratory; Bergfors (1999) Protein Crystallization International University Line; Mullin (1993) Crystallization Butterwoth-Heinemann; Stout and Jensen (1989) X-ray structure determination: a practical guide, 2nd Edition Wiley Publishers, New York; Ladd and Palmer (1993) Structure determination by X-ray crystallography, 3rd Edition Plenum Press, NewYork; Blundell and Johnson (1976) Protein Crystallography Academic Press, New York; Glusker and Trueblood (1985) Crystal structure analysis: A primer, 2nd Ed. Oxford University Press, NewYork; International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules; McPherson (2002) Introduction to Macromolecular Crystallography Wiley-Liss; McRee and David (1999) Practical Protein Crystallography, Second Edition Academic Press; Drenth (1999) Principles of Protein X-Ray Crystallography (Springer Advanced Texts in

Chemistry) Springer-Verlag; Fanchon and Hendrickson (1991) Chapter 15 of Crystallographic Computing, Volume 5 IUCr/Oxford University Press; Murthy (1996) Chapter 5 of Crystallographic Methods and Protocols Humana Press; Dauter et al. (2000) "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides" Acta Cryst.D56:232-237; Dauter (2002) "New approaches to high-throughput phasing" Curr. Opin. Structural Biol. 12:674–678; Chen et al. (1991) "Crystal structure of a bovine neurophysin-II dipeptide complex at 2.8 Å determined from the single-wavelength anomalous scattering signal of an incorporated iodine atom" Proc. Natl Acad. Sci. USA, 88:4240–4244; and Gavira et al. (2002) "*Ab initio* crystallographic structure determination of insulin from protein to electron density without crystal handling" Acta Cryst.D58:1147-1154.

[0208] In addition, a variety of programs to facilitate data collection, phase determination, model building and refinement, and the like are publicly available. Examples include, but are not limited to, the HKL2000 package (Otwinowski and Minor (1997) "Processing of X-ray Diffraction Data Collected in Oscillation Mode" Methods in Enzymology 276:307-326), the CCP4 package (Collaborative Computational Project (1994) "The CCP4 suite: programs for protein crystallography" Acta Crystallogr D 50:760-763), SOLVE and RESOLVE (Terwilliger and Berendzen (1999) Acta Crystallogr D 55 (Pt 4):849-861), SHELXS and SHELXD (Schneider and Sheldrick (2002) "Substructure solution with SHELXD" Acta Crystallogr D Biol Crystallogr 58:1772-1779), Refmac5 (Murshudov et al. (1997) "Refinement of Macromolecular Structures by the Maximum-Likelihood Method" Acta Crystallogr D 53:240-255), PRODRG (van Aalten et al. (1996) "PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules" J Comput Aided Mol Des 10:255-262), and O (Jones et al. (1991) "Improved methods for building protein models in electron density maps and the location of errors in these models" Acta Crystallogr A 47 (Pt 2):110-119).

[0209] Techniques for structure determination by NMR spectroscopy are similarly well described in the literature. See, e.g., Cavanagh et al. (1995) Protein NMR Spectroscopy: Principles and Practice, Academic Press; Levitt (2001) Spin Dynamics: Basics of Nuclear Magnetic Resonance, John Wiley & Sons; Evans (1995) Biomolecular NMR Spectroscopy, Oxford University Press; Wüthrich (1986) NMR of Proteins and Nucleic Acids (Baker Lecture

Series), Kurt Wiley-Interscience; Neuhaus and Williamson (2000) The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd Edition, Wiley-VCH; Macomber (1998) A Complete Introduction to Modern NMR Spectroscopy, Wiley-Interscience; Downing (2004) Protein NMR Techniques (Methods in Molecular Biology), 2nd edition, Humana Press; Clore and Gronenborn (1994) NMR of Proteins (Topics in Molecular and Structural Biology), CRC Press; Reid (1997) Protein NMR Techniques, Humana Press; Krishna and Berliner (2003) Protein NMR for the Millenium (Biological Magnetic Resonance), Kluwer Academic Publishers; Kiihne and De Groot (2001) Perspectives on Solid State NMR in Biology (Focus on Structural Biology, 1), Kluwer Academic Publishers; Jones et al. (1993) Spectroscopic Methods and Analyses: NMR, Mass Spectrometry, and Related Techniques (Methods in Molecular Biology, Vol. 17), Humana Press; Goto and Kay (2000) Curr. Opin. Struct. Biol. 10:585; Gardner (1998) Annu. Rev. Biophys. Biomol. Struct. 27:357; Wüthrich (2003) Angew. Chem. Int. Ed. 42:3340; Bax (1994) Curr. Opin. Struct. Biol. 4:738; Pervushin et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12366; Fiaux et al. (2002) Nature 418:207; Fernandez and Wider (2003) Curr. Opin. Struct. Biol. 13:570; Ellman et al. (1992) J. Am. Chem. Soc. 114:7959; Wider (2000) BioTechniques 29:1278-1294; Pellecchia et al. (2002) Nature Rev. Drug Discov. (2002) 1:211-219; Arora and Tamm (2001) Curr. Opin. Struct. Biol. 11:540-547; Flaux et al. (2002) Nature 418:207-211; Pellecchia et al. (2001) J. Am. Chem. Soc. 123:4633-4634; and Pervushin et al. (1997) Proc. Natl. Acad. Sci. USA 94:12366-12371.

[0210] The structure of a polymerase, or polymerase bound to a DNA or with a given nucleotide analog incorporated into the active site can, as noted, be directly determined, e.g., by x-ray crystallography or NMR spectroscopy, or the structure can be modeled based on the structure of the polymerase and/or a structure of a polymerase with a natural nucleotide bound. The active site or other relevant domain of the polymerase can be identified, for example, by homology with other polymerases, examination of polymerase-template or polymerase-nucleotide co-complexes, biochemical analysis of mutant polymerases, and/or the like. The position of a nucleotide analog (as opposed to an available nucleotide structure) in the active site can be modeled, for example, by projecting the location of non-natural features of the analog (e.g., additional phosphate or phosphonate groups in the phosphorus containing chain linked to the nucleotide, e.g., tetra, penta or hexa phosphate groups, detectable labeling groups,

e.g., fluorescent dyes, or the like) based on the previously determined location of another nucleotide or nucleotide analog in the active site.

[0211] Such modeling of the nucleotide analog or template (or both) in the active site can involve simple visual inspection of a model of the polymerase, for example, using molecular graphics software such as the PyMOL viewer (open source, freely available on the World Wide Web at www(dot)pymol(dot)org) or Insight II (commercially available from Accelrys at (www (dot) accelrys (dot) com/products/insight). Alternatively, modeling of the active site complex of the polymerase or a putative mutant polymerase, for example, can involve computer-assisted docking, molecular dynamics, free energy minimization, and/or like calculations. Such modeling techniques have been well described in the literature; see, e.g., Babine and Abdel-Meguid (eds.) (2004) Protein Crystallography in Drug Design, Wiley-VCH, Weinheim; Lyne (2002) "Structure-based virtual screening: An overview" Drug Discov. Today 7:1047-1055; Molecular Modeling for Beginners, at (www (dot) usm (dot) maine (dot) edu/~rhodes/SPVTut/index (dot) html; and Methods for Protein Simulations and Drug Design at (www (dot) dddc (dot) ac (dot) cn/embo04; and references therein. Software to facilitate such modeling is widely available, for example, the CHARMm simulation package, available academically from Harvard University or commercially from Accelrys (at www (dot) accelrys (dot) com), the Discover simulation package (included in Insight II, *supra*), and Dynama (available at (www(dot) cs (dot) gsu (dot) edu/~cscrwh/progs/progs (dot) html). See also an extensive list of modeling software at (www (dot) netsci (dot) org/Resources/Software/Modeling/MMMD/top (dot) html.

[0212] Visual inspection and/or computational analysis of a polymerase model can identify relevant features of the active site or other domain, including, for example, amino acid residues of domains that are in close proximity to one another (to stabilize inter-domain interactions) residues in the active site that interact with the nucleotide or analog, or that modulate how large a binding pocket for the analog is relative to the analog. That is, inter-domain amino acid contacts can stabilize the closed complex, and/or the size or composition (e.g., position of charged or hydrophobic residues) of the binding pocket in the active site can control entry and release of the nucleotide, which can affect branching rate. A residue can, for example, be deleted or replaced with a residue having a different (smaller, larger, ionic, non-

ionic, etc.) side chain. Similarly, residues that can be altered to introduce desirable interactions with the nucleotide analog can be identified to reduce branching. Such a residue can be replaced with a residue that is complementary with, e.g., a non-natural feature of the analog, for example, with a residue that can hydrogen bond to the analog (e.g., serine, threonine, histidine, asparagine, or glutamine), a hydrophobic residue that can interact with a hydrophobic group on the analog, an aromatic residue that can provide favorable hydrophobic interactions with a group on the analog (e.g., a fluorophore), an aromatic residue that can engage in a π-π or edge-face stacking interaction with an aromatic group in the analog, a residue that can engage in a cation-π interaction with the analog, or a charged residue (e.g., aspartic or glutamic acid, or lysine, arginine, or histidine) that can electrostatically interact with an oppositely charged moiety on the analog (e.g., an additional phosphate group).

[0213] Thus, in addition to methods of using the polymerases and other compositions herein, the present invention also includes methods of making the polymerases. As described, methods of making a recombinant DNA polymerase can include structurally modeling a first polymerase, e.g., using any available crystal structure and molecular modeling software or system. Based on the modeling, one or more feature affecting closed complex stability, or nucleotide access or removal to or from the active site (and, thereby, branching) and/or binding of a DNA or nucleotide analog within the active site region is identified. These residues can be, e.g., in the active site, an exonuclease, TPR2 or thumb domain (or interface between domains) or proximal to such domains. The DNA polymerase is mutated to include non-natural residues at such positions, and then screened for an activity of interest.

Mutating Polymerases

[0214] Various types of mutagenesis are optionally used in the present invention, e.g., to modify polymerases to produce variants, e.g., in accordance with polymerase models and model predictions as discussed above, or using random or semi-random mutational approaches. In general, any available mutagenesis procedure can be used for making polymerase mutants. Such mutagenesis procedures optionally include selection of mutant nucleic acids and polypeptides for one or more activity of interest (e.g., decreased branch fraction, increased or decreased complex stability, improved processivity, and/or improved k_{off}, K_m, V_{max}, k_{cat} etc.,

e.g., for a given nucleotide analog). Procedures that can be used include, but are not limited to: site-directed point mutagenesis, random point mutagenesis, in vitro or in vivo homologous recombination (DNA shuffling and combinatorial overlap PCR), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA, point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, degenerate PCR, double-strand break repair, and many others known to persons of skill. The starting polymerase for mutation can be any of those noted herein, including available polymerase mutants such as those identified e.g., in WO 2007/076057 POLYMERASES FOR NUCLEOTIDE ANALOG INCORPORATION by Hanzel et al.; PCT/US2007/022459 POLYMERASE ENZYMES AND REAGENTS FOR ENHANCED NUCLEIC ACID SEQUENCING; Hanzel et al. WO 2007/075987 ACTIVE SURFACE COUPLED POLYMERASES; and Hanzel et al. WO 2007/076057 PROTEIN ENGINEERING STRATEGIES TO OPTIMIZE ACTIVITY OF SURFACE ATTACHED PROTEINS.

[0215] Optionally, mutagenesis can be guided by known information from a naturally occurring polymerase molecule, or of a known altered or mutated polymerase (e.g., using an existing mutant polymerase as noted in the preceding references), e.g., sequence, sequence comparisons, physical properties, crystal structure and/or the like as discussed above. However, in another class of embodiments, modification can be essentially random (e.g., as in classical or "family" DNA shuffling, see, e.g., Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288–291).

[0216] Additional information on mutation formats is found in: Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2000 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2006) ("Ausubel")) and PCR Protocols A Guide to Methods and Applications (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis). The following publications and references cited within provide additional detail on mutation formats: Arnold, *Protein engineering for unusual environments*,

Current Opinion in Biotechnology 4:450-455 (1993); Bass et al., *Mutant Trp repressors with new DNA-binding specificities*, Science 242:240-245 (1988); Bordo and Argos (1991) *Suggestions for "safe Residue Substitutions in Site-directed Mutagenesis* 217:721-729; Botstein & Shortle, *Strategies and applications of in vitro mutagenesis*, Science 229:1193-1201(1985); Carter et al., *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, Nucl. Acids Res. 13: 4431-4443 (1985); Carter, *Site-directed mutagenesis*, Biochem. J. 237:1-7 (1986); Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*, Methods in Enzymol. 154: 382-403 (1987); Dale et al., *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, Methods Mol. Biol. 57:369-374 (1996); Eghtedarzadeh & Henikoff, *Use of oligonucleotides to generate large deletions*, Nucl. Acids Res. 14: 5115 (1986); Fritz et al., *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, Nucl. Acids Res. 16: 6987-6999 (1988); Grundström et al., *Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis*, Nucl. Acids Res. 13: 3305-3316 (1985); Hayes (2002) *Combining Computational and Experimental Screening for rapid Optimization of Protein Properties* PNAS 99(25) 15926-15931; Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin)) (1987); Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel et al., *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Methods in Enzymol. 154, 367-382 (1987); Kramer et al., *The gapped duplex DNA approach to oligonucleotide-directed mutation construction*, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer & Fritz *Oligonucleotide-directed construction of mutations via gapped duplex DNA*, Methods in Enzymol. 154:350-367 (1987); Kramer et al., *Point Mismatch Repair*, Cell 38:879-887 (1984); Kramer et al., *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, Nucl. Acids Res. 16: 7207 (1988); Ling et al., *Approaches to DNA mutagenesis: an overview*, Anal Biochem. 254(2): 157-178 (1997); Lorimer and Pastan Nucleic Acids Res. 23, 3067-8 (1995); Mandecki, *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Nakamaye & Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by*

phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, Nucl. Acids Res. 14: 9679-9698 (1986); Nambiar et al., *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, Science 223: 1299-1301 (1984); Sakamar and Khorana, *Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, Nucl. Acids Res. 14: 6361-6372 (1988); Sayers et al., *Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 16:791-802 (1988); Sayers et al., *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) Nucl. Acids Res. 16: 803-814; Sieber, et al., Nature Biotechnology, 19:456-460 (2001); Smith, *In vitro mutagenesis*, Ann. Rev. Genet. 19:423-462(1985); Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Stemmer, Nature 370, 389-91 (1994); Taylor et al., *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor et al., *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, Nucl. Acids Res. 13: 8765-8787 (1985); Wells et al., *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986); Wells et al., *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*, Gene 34:315-323 (1985); Zoller & Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment*, Nucleic Acids Res. 10:6487-6500 (1982); Zoller & Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, Methods in Enzymol. 100:468-500 (1983); Zoller & Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template*, Methods in Enzymol. 154:329-350 (1987); Clackson et al. (1991) "Making antibody fragments using phage display libraries" Nature 352:624-628; Gibbs et al. (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" Gene 271:13-20; and Hiraga and Arnold (2003) "General method for sequence-independent site-directed chimeragenesis: J. Mol. Biol. 330:287-296. Additional details on many of the above methods can be found in Methods in

Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Determining Kinetic Parameters

[0217] The polymerases of the invention can be screened or otherwise tested to determine whether the polymerase displays a modified activity for or with a nucleotide analog or template as compared to the first DNA polymerase (e.g., a corresponding wild-type or available mutant polymerase from which the recombinant polymerase of the invention was derived). For example, branch fraction, k_{off}, k_{cat}, K_m, V_{max}, k_{cat}/K_m, V_{max}/K_m, k_{pol}, and/or K_d of the recombinant DNA polymerase for the nucleotide (or analog) or template nucleic acid can be determined. The enzyme perfection metric k_{cat}/K_m is also a useful measure, e.g., for assessing branch rate. k_{cat}/K_m is a measure of substrate binding that leads to product formation (and, thus, includes terms defining binding K_d and inversely predicts branching fraction formation).

[0218] As is well-known in the art, for enzymes obeying simple Michaelis-Menten kinetics, kinetic parameters are readily derived from rates of catalysis measured at different substrate concentrations. The Michaelis-Menten equation, $V=V_{max}[S]([S]+K_m)^{-1}$, relates the concentration of uncombined substrate ([S], approximated by the total substrate concentration), the maximal rate (V_{max}, attained when the enzyme is saturated with substrate), and the Michaelis constant (K_m, equal to the substrate concentration at which the reaction rate is half of its maximal value), to the reaction rate (V). Details regarding K_{off} determination are described above. In general, the dissociation rate can be measured in any manner that detects the polymerase/ DNA complex over time. This includes stopped flow spectroscopy, or even simply by taking aliquots over time and testing for polymerase activity on the template of interest. Free polymerase is captured with a polymerase trap after dissociation, e.g., by incubation in the presence of heparin or an excess of competitor DNA (e.g., non-specific salmon sperm DNA, or the like).

[0219] For many enzymes, K_m is equal to the dissociation constant of the enzyme-substrate complex and is thus a measure of the strength of the enzyme-substrate complex. For such an enzyme, in a comparison of K_ms, a lower K_m represents a complex with stronger binding, while a higher Km represents a complex with weaker binding. The ratio k_{cat}/K_m,

sometimes called the specificity constant, represents the apparent rate constant for combination of substrate with free enzyme. The larger the specificity constant, the more efficient the enzyme is in binding the substrate and converting it to product (this provides an inverse measure of branching rate, as branching rate is the rate at which the enzyme binds substrate (e.g., nucleotide), but does not convert it to product (e.g., a DNA polymer).

[0220] k_{cat} (also called the turnover number of the enzyme) can be determined if the total enzyme concentration ($[E_T]$, i.e., the concentration of active sites) is known, since $V_{max}=k_{cat}[E_T]$. For situations in which the total enzyme concentration is difficult to measure, the ratio V_{max}/K_m is often used instead as a measure of efficiency. K_m and V_{max} can be determined, for example, from a Lineweaver-Burk plot of 1/V against 1/[S], where the y intercept represents $1/V_{max}$, the x intercept $-1/K_m$, and the slope K_m/V_{max}, or from an Eadie-Hofstee plot of V against V/[S], where the y intercept represents V_{max}, the x intercept V_{max}/K_m, and the slope $-K_m$. Software packages such as KinetAsyst™ or Enzfit (Biosoft, Cambridge, UK) can facilitate the determination of kinetic parameters from catalytic rate data.

[0221] For enzymes such as polymerases that have multiple substrates, varying the concentration of only one substrate while holding the others in suitable excess (e.g., effectively constant) concentration typically yields normal Michaelis-Menten kinetics.

[0222] In one embodiment, using pre-steady-state kinetics, the nucleotide concentration dependence of the rate k_{obs} (the observed first-order rate constant for dNTP incorporation) provides an estimate of the K_m for a ground state binding and the maximum rate of polymerization (k_{pol}). The k_{obs} is measured using a burst assay. The results of the assay are fitted with the Burst equation; Product = $A[1-\exp(-k_{obs}*t)]+k_{ss}*t$ where A represents amplitude an estimate of the concentration of the enzyme active site*s, k_{ss} is the observed steady-state rate constant and t is the reaction incubation time. The K_m for dNTP binding to the polymerase-DNA complex and the k_{pol} are calculated by fitting the dNTP concentration dependent change in the k_{obs} using the equation $k_{obs} = (k_{pol}*[S])*(K_m+[S])-1$ where [S] is the substrate concentration. Results are optionally obtained from a rapid-quench experiment (also called a quench-flow measurement), for example, based on the methods described in Johnson (1986) "Rapid kinetic analysis of mechanochemical adenosinetriphosphatases" Methods Enzymol.

134:677-705, Patel et al. (1991) "Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant" Biochemistry 30(2):511-25, and Tsai and Johnson (2006) "A new paradigm for DNA polymerase specificity" Biochemistry 45(32):9675-87.

[0223] Parameters such as rate of binding of a nucleotide analog or template by the recombinant polymerase, rate of product release by the recombinant polymerase, or branching rate of the recombinant polymerase can also be determined, and optionally compared to that of the first polymerase (e.g., a corresponding wild-type polymerase).

[0224] For a more thorough discussion of enzyme kinetics, see, e.g., Berg, Tymoczko, and Stryer (2002) Biochemistry, Fifth Edition, W. H. Freeman; Creighton (1984) Proteins: Structures and Molecular Principles, W. H. Freeman; and Fersht (1985) Enzyme Structure and Mechanism, Second Edition, W. H. Freeman.

[0225] In one aspect, the improved property of the enzymes of the invention is measured with reference to a model analog or analog set and compared with a given parental enzyme. For example, in the case of enzymes derived from a Φ29 parental enzyme, where the improvement being sought is a decrease in stability of the closed complex, an improved enzyme of the invention (i.e., an enzyme with an increased branching fraction) would have a higher k_{off} than the parental enzyme, e.g., wild type Φ29. While the foregoing may be used as a characterization tool, it in no way is intended as a specifically limiting reaction of the invention.

Screening Polymerases

[0226] Screening or other protocols can be used to determine whether a polymerase displays a modified activity for a nucleotide analog as compared to the first DNA polymerase. For example, k_{off}, k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for the template or nucleotide or analog can be determined as discussed above.

[0227] In one desirable aspect, a library of recombinant DNA polymerases can be made and screened for these properties. For example, a plurality of members of the library can be made to include one or more mutations that increase branching fractions, delay translocation

and/or increase residence time, that are then screened for the properties of interest (e.g., increased branching fraction, delayed translocation or increased residence time). In general, the library can be screened to identify at least one member comprising a modified activity of interest.

[0228] Libraries of polymerases can be either physical or logical in nature. Moreover, any of a wide variety of library formats can be used. For example, polymerases can be fixed to solid surfaces in arrays of proteins. Similarly, liquid phase arrays of polymerases (e.g., in microwell plates) can be constructed for convenient high-throughput fluid manipulations of solutions comprising polymerases. Liquid, emulsion, or gel-phase libraries of cells that express recombinant polymerases can also be constructed, e.g., in microwell plates, or on agar plates. Phage display libraries of polymerases or polymerase domains (e.g., including the active site region or interdomain stability regions) can be produced. Likewise, yeast display libraries can be used. Instructions in making and using libraries can be found, e.g., in Sambrook, Ausubel and Berger, referenced herein.

[0229] For the generation of libraries involving fluid transfer to or from microtiter plates, a fluid handling station is optionally used. Several "off the shelf" fluid handling stations for performing such transfers are commercially available, including e.g., the Zymate systems from Caliper Life Sciences (Hopkinton, MA) and other stations which utilize automatic pipettors, e.g., in conjunction with the robotics for plate movement (e.g., the ORCA® robot, which is used in a variety of laboratory systems available, e.g., from Beckman Coulter, Inc. (Fullerton, CA).

[0230] In an alternate embodiment, fluid handling is performed in microchips, e.g., involving transfer of materials from microwell plates or other wells through microchannels on the chips to destination sites (microchannel regions, wells, chambers or the like). Commercially available microfluidic systems include those from Hewlett-Packard/Agilent Technologies (e.g., the HP2100 bioanalyzer) and the Caliper High Throughput Screening System. The Caliper High Throughput Screening System provides one example interface between standard microwell library formats and Labchip technologies. RainDance Technologies' nanodroplet platform provides another method for handling large numbers of