PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Michael		Schmitt	

Residence Information (Select One) ● US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Seattle	WA	US

Mailing Address of Inventor:

Address 1	4311 11th Avenue NE
Address 2	
City	Seattle
State/Province	WA
Postal Code	98105-4608
Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Jesse		Salk	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Seattle	WA	US

Mailing Address of Inventor:

Address 1	4311 11th Avenue NE
Address 2	
City	Seattle
State/Province	WA
Postal Code	98105-4608
Country	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Lawrence	A.	Loeb	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

City	Bellevue	State/Province	WA	Country of Residence	US

Mailing Address of Inventor:

Address 1	4311 11th Avenue NE		
Address 2			
City	Seattle	**State/Province**	WA
Postal Code	98105-4608	**Country**	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	94991		
Email Address	patentprocurement@perkinscoie.com	Add Email	Remove Email

Application Information:

Title of the Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		
Attorney Docket Number	072227-8043.US07	**Small Entity Status Claimed**	☒
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	12	**Suggested Figure for Publication (if any)**	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	● Customer Number	US Patent Practitioner	◯ Limited Recognition (37 CFR 11.9)
Customer Number	94991		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Continuation of	15660785	2017-07-26
Prior Application Status	Patented		Remove

Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)	Patent Number	Issue Date (YYYY-MM-DD)
15660785	Continuation of	14386800	2014-09-20	9752188	2017-09-05

Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
14386800	a 371 of international	PCT/US2013/032665	2013-03-15
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCT/US2013/032665	Claims benefit of provisional	61613413	2012-03-20
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCT/US2013/032665	Claims benefit of provisional	61625623	2012-04-17

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
PCT/US2013/032665	Claims benefit of provisional	61625319	2012-04-17
Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button.			Add

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

			Remove
Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)
Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.			Add

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

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

Authorization or Opt-Out of Authorization to Permit Access:

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

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

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

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

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

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

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

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

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

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

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

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1 Remove

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

Assignee	Legal Representative under 35 U.S.C. 117	Joint Inventor
● Person to whom the inventor is obligated to assign.	Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:

If the Applicant is an Organization check here. ☒

Organization Name	UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION		
Mailing Address Information For Applicant:			
Address 1	4311 11th Avenue NE		
Address 2			
City	Seattle	**State/Province**	WA
Country	US	Postal Code	98105-4608
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button. Add

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

Remove

If the Assignee or Non-Applicant Assignee is an Organization check here. ☐

Prefix	**Given Name**	Middle Name	**Family Name**	Suffix

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1			
Address 2			
City		**State/Province**	
Country i		Postal Code	
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add

Signature:

Remove

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

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

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

Signature	/Yang Tang/			Date (YYYY-MM-DD)	2018-08-30
First Name	Yang	Last Name	Tang	Registration Number	55663

Additional Signature may be generated within this form by selecting the Add button. Add

PTO/AIA/14 (11-15)
Approved for use through 04/30/2017. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	072227-8043.US07
	Application Number	
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

This collection of information is required by 37 CFR 1.76. 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 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet 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.**

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 the Freedom of Information Act requires disclosure of these records.

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 CooperationTreaty.

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 inspections 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.

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

PRIORITY CLAIM

[0001] This application claims priority to U.S. Provisional Patent Application No. 61/613,413, filed March 20, 2012; U.S. Provisional Patent Application No. 61/625,623, filed April 17, 2012; and U.S. Provisional Patent Application No. 61/625,319, filed April 17, 2012; the subject matter of all of which are hereby incorporated by reference as if fully set forth herein.

STATEMENT OF GOVERNMENT INTEREST

[0002] The present invention was made with government support under Grant Nos. RO1 CA115802 and RO1 CA102029 awarded by the National Institutes of Health. The Government has certain rights in the invention.

BACKGROUND

[0003] The advent of massively parallel DNA sequencing has ushered in a new era of genomic exploration by making simultaneous genotyping of hundreds of billions of base-pairs possible at small fraction of the time and cost of traditional Sanger methods [1]. Because these technologies digitally tabulate the sequence of many individual DNA fragments, unlike conventional techniques which simply report the average genotype of an aggregate collection of molecules, they offer the unique ability to detect minor variants within heterogeneous mixtures [2].

[0004] This concept of "deep sequencing" has been implemented in a variety fields including metagenomics [3, 4], paleogenomics [5], forensics [6], and human genetics [7, 8] to disentangle subpopulations in complex biological samples. Clinical applications, such prenatal screening for fetal aneuploidy [9, 10], early detection of cancer [11] and monitoring its response to therapy [12, 13] with nucleic acid-based serum biomarkers, are rapidly being developed. Exceptional diversity within microbial [14, 15] viral [16-18] and tumor cell populations [19, 20] has been characterized through next-generation sequencing, and many low-frequency, drug-resistant variants of

therapeutic importance have been so identified [12, 21, 22]. Previously unappreciated intra-organismal mosasism in both the nuclear [23] and mitochondrial [24, 25] genome has been revealed by these technologies, and such somatic heterogeneity, along with that arising within the adaptive immune system [13], may be an important factor in phenotypic variability of disease.

[0005] Deep sequencing, however, has limitations. Although, in theory, DNA subpopulations of any size should be detectable when deep sequencing a sufficient number of molecules, a practical limit of detection is imposed by errors introduced during sample preparation and sequencing. PCR amplification of heterogeneous mixtures can result in population skewing due to stoichastic and non-stoichastic amplification biases and lead to over- or under-representation of particular variants [26]. Polymerase mistakes during pre-amplification generate point mutations resulting from base mis-incorporations and rearrangements due to template switching [26, 27]. Combined with the additional errors that arise during cluster amplification, cycle sequencing and image analysis, approximately 1% of bases are incorrectly identified, depending on the specific platform and sequence context [2, 28]. This background level of artifactual heterogeneity establishes a limit below which the presence of true rare variants is obscured [29].

[0006] A variety of improvements at the level of biochemistry [30-32] and data processing [19, 21, 28, 32, 33] have been developed to improve sequencing accuracy. The ability to resolve subpopulations below 0.1%, however, has remained elusive. Although several groups have attempted to increase sensitivity of sequencing, several limitations remain. For example techniques whereby DNA fragments to be sequenced are each uniquely tagged [34, 35] prior to amplification [36-41] have been reported. Because all amplicons derived from a particular starting molecule will bear its specific tag, any variation in the sequence or copy number of identically tagged sequencing reads can be discounted as technical error. This approach has been used to improve counting accuracy of DNA [38, 39, 41] and RNA templates [37, 38, 40] and to correct base errors arising during PCR or sequencing [36, 37, 39]. Kinde et. al. reported a reduction in error frequency of approximately 20-fold with a tagging method that is based on labeling single-stranded DNA fragments with a primer containing a 14 bp

degenerate sequence. This allowed for an observed mutation frequency of ~0.001% mutations/bp in normal human genomic DNA [36]. Nevertheless, a number of highly sensitive genetic assays have indicated that the true mutation frequency in normal cells is likely to be far lower, with estimates of per-nucleotide mutation frequencies generally ranging from 10^{-9} to 10^{-11} [42]. Thus, the mutations seen in normal human genomic DNA by Kinde et al. are likely the result of significant technical artifacts.

[0007] Traditionally, next-generation sequencing platforms rely upon generation of sequence data from a single strand of DNA. As a consequence, artifactual mutations introduced during the initial rounds of PCR amplification are undetectable as errors - even with tagging techniques - if the base change is propagated to all subsequent PCR duplicates. Several types of DNA damage are highly mutagenic and may lead to this scenario. Spontaneous DNA damage arising from normal metabolic processes results in thousands of damaging events per cell per day [43]. In addition to damage from oxidative cellular processes, further DNA damage is generated *ex vivo* during tissue processing and DNA extraction [44]. These damage events can result in frequent copying errors by DNA polymerases: for example a common DNA lesion arising from oxidative damage, 8-oxo-guanine, has the propensity to incorrectly pair with adenine during complementary strand extension with an overall efficiency greater than that of correct pairing with cytosine, and thus can contribute a large frequency of artifactual G→T mutations [45]. Likewise, deamination of cytosine to form uracil is a particularly common event which leads to the inappropriate insertion of adenine during PCR, thus producing artifactual C→T mutations with a frequency approaching 100% [46].

[0008] It would be desirable to develop an approach for tag-based error correction, which reduces or eliminates artifactual mutations arising from DNA damage, PCR errors, and sequencing errors; allows rare variants in heterogeneous populations to be detected with unprecedented sensitivity; and which capitalizes on the redundant information stored in complexed double-stranded DNA.

SUMMARY

[0009] In one embodiment, a single molecule identifier (SMI) adaptor molecule for use in sequencing a double-stranded target nucleic acid molecule is provided. Said

SMI adaptor molecule includes a single molecule identifier (SMI) sequence which comprises a degenerate or semi-degenerate DNA sequence; and an SMI ligation adaptor that allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence. The SMI sequence may be single-stranded or double-stranded. In some embodiments, the double-stranded target nucleic acid molecule is a double-stranded DNA or RNA molecule.

[0010] In another embodiment, a method of obtaining the sequence of a double-stranded target nucleic acid is provided (also known as Duplex Consensus Sequencing or DCS) is provided. Such a method may include steps of ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex; amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products; and sequencing the amplified SMI-target nucleic acid products.

[0011] In some embodiments, the method may additionally include generating an error-corrected double-stranded consensus sequence by (i) grouping the sequenced SMI-target nucleic acid products into families of paired target nucleic acid strands based on a common set of SMI sequences; and (ii) removing paired target nucleic acid strands having one or more nucleotide positions where the paired target nucleic acid strands are non-complementary (or alternatively removing individual nucleotide positions in cases where the sequence at the nucleotide position under consideration disagrees among the two strands). In further embodiments, the method confirms the presence of a true mutation by (i) identifying a mutation present in the paired target nucleic acid strands having one or more nucleotide positions that disagree; (ii) comparing the mutation present in the paired target nucleic acid strands to the error corrected double-stranded consensus sequence; and (iii) confirming the presence of a true mutation when the mutation is present on both of the target nucleic acid strands and appears in all members of a paired target nucleic acid family.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 illustrates an overview of Duplex Consensus Sequencing. Sheared double-stranded DNA that has been end-repaired and T-tailed is combined

with A-tailed SMI adaptors and ligated according to one embodiment. Because every adaptor contains a unique, double-stranded, complementary n-mer random tag on each end (n-mer = 12 bp according to one embodiment), every DNA fragment becomes labeled with two distinct SMI sequences (arbitrarily designated α and β in the single capture event shown). After size-selecting for appropriate length fragments, PCR amplification with primers containing Illumina flow-cell-compatible tails is carried out to generate families of PCR duplicates. By virtue of the asymmetric nature of adapted fragments, two types of PCR products are produced from each capture event. Those derived from one strand will have the α SMI sequence adjacent to flow-cell sequence 1 and the β SMI sequence adjacent to flow cell sequence 2. PCR products originating from the complementary strand are labeled reciprocally.

[0013] Figure 2 illustrates Single Molecule Identifier (SMI) adaptor synthesis according to one embodiment. Oligonucleotides are annealed and the complement of the degenerate lower arm sequence (N's) plus adjacent fixed bases is produced by polymerase extension of the upper strand in the presence of all four dNTPs. After reaction cleanup, complete adaptor A-tailing is ensured by extended incubation with polymerase and dATP.

[0014] Figure 3 illustrates error correction through Duplex Consensus Sequencing (DCS) analysis according to one embodiment. (a-c) shows sequence reads (brown) sharing a unique set of SMI tags are grouped into paired families with members having strand identifiers in either the αβ or βα orientation. Each family pair reflects one double-stranded DNA fragment. (a) shows mutations (spots) present in only one or a few family members representing sequencing mistakes or PCR-introduced errors occurring late in amplification. (b) shows mutations occurring in many or all members of one family in a pair representing mutations scored on only one of the two strands, which can be due to PCR errors arising during the first round of amplification such as might occur when copying across sites of mutagenic DNA damage. (c) shows true mutations (* arrow) present on both strands of a captured fragment appear in all members of a family pair. While artifactual mutations may co-occur in a family pair with a true mutation, these can be independently identified and discounted when producing (d) an error-corrected consensus sequence (i.e., single stranded consensus sequence)

(+ arrow) for each duplex. (e) shows consensus sequences from all independently captured, randomly sheared fragments containing a particular genomic site are identified and (f) compared to determine the frequency of genetic variants at this locus within the sampled population.

[0015] Figure 4 illustrates an example of how a SMI sequence with n-mers of 4 nucleotides in length (4-mers) are read by Duplex Consensus Sequencing (DCS) according to some embodiments. (A) shows the 4-mers with the PCR primer binding sites (or flow cell sequences) 1 and 2 indicated at each end. (B) shows the same molecules as in (A) but with the strands separated and the lower strand now written in the 5'-3' direction. When these molecules are amplified with PCR and sequenced, they will yield the following sequence reads: The top strand will give a read 1 file of TAAC--- and a read 2 file of GCCA---. Combining the read 1 and read 2 tags will give TAACCGGA as the SMI for the top strand. The bottom strand will give a read 1 file of CGGA---- and a read 2 file of TAAC---. Combining the read 1 and read 2 tags will give CGGATAAC as the SMI for the bottom strand. (C) illustrates the orientation of paired strand mutations in DCS. In the initial DNA duplex shown in Figures 4A and 4B, a mutation "x" (which is paired to a complementary nucleotide "y") is shown on the left side of the DNA duplex. The "x" will appear in read 1, and the complementary mutation on the opposite strand, "y," will appear in read 2. Specifically, this would appear as "x" in both read 1 and read 2 data, because "y" in read 2 is read out as "x" by the sequencer owing to the nature of the sequencing primers, which generate the complementary sequence during read 2.

[0016] Figure 5 illustrates duplex sequencing of human mitochondrial DNA. (A) Overall mutation frequency as measured by a standard sequencing approach, SSCS, and DCS. (B) Pattern of mutation in human mitochondrial DNA by a standard sequencing approach. The mutation frequency (vertical axis) is plotted for every position in the ~16-kb mitochondrial genome. Due to the substantial background of technical error, no obvious mutational pattern is discernible by this method. (C) DCS analysis eliminates sequencing artifacts and reveals the true distribution of mitochondrial mutations to include a striking excess adjacent to the mtDNA origin of replication. (D)

SSCS analysis yields a large excess of G→T mutations relative to complementary C→A mutations, consistent with artifacts from damaged-induced 8-oxo-G lesions during PCR. All significant ($P < 0.05$) differences between paired reciprocal mutation frequencies are noted. (E) DCS analysis removes the SSCS strand bias and reveals the true mtDNA mutational spectrum to be characterized by an excess of transitions.

[0017] Figure 6 shows that consensus sequencing removes artifactual sequencing errors as compared to Raw Reads. Duplex Consensus Sequencing (DCS) results in an approximately equal number of mutations as the reference and single strand consensus sequencing (SSCS) .

[0018] Figure 7 illustrates duplex sequencing of M13mp2 DNA. (A) Single-strand consensus sequences (SSCSs) reveal a large excess of G→A/C→T and G→T/C→A mutations, whereas duplex consensus sequences (DCSs) yield a balanced spectrum. Mutation frequencies are grouped into reciprocal mispairs, as DCS analysis only scores mutations present in both strands of duplex DNA. All significant ($P < 0.05$) differences between DCS analysis and the literature reference values are noted. (B) Complementary types of mutations should occur at approximately equal frequencies within a DNA fragment population derived from duplex molecules. However, SSCS analysis yields a 15-fold excess of G→T mutations relative to C→A mutations and an 11-fold excess of C→T mutations relative to G→A mutations. All significant ($P < 0.05$) differences between paired reciprocal mutation frequencies are noted.

[0019] Figure 8 shows the effect of DNA damage on the mutation spectrum. DNA damage was induced by incubating purified M13mp2 DNA with hydrogen peroxide and FeSO4. (A) SSCS analysis reveals a further elevation from baseline of G→T mutations, indicating these events to be the artifactual consequence of nucleotide oxidation. All significant ($P < 0.05$) changes from baseline mutation frequencies are noted. (B) Induced DNA damage had no effect on the overall frequency or spectrum of DCS mutations.

[0020] Figure 9 shows duplex sequencing results in accurate recovery of spiked-control mutations. A series of variants of M13mp2 DNA, each harboring a known single-nucleotide substitution, were mixed in together at known ratios and the mixture

was sequenced to ~20,000-fold final depth. Standard sequencing analysis cannot accurately distinguish mutants present at a ratio of less than 1/100, because artifactural mutations occurring at every position obscure the presence of less abundant true mutations, rendering apparent recovery greater than 100%. Duplex consensus sequences, in contrast, accurately identify spiked-in mutations down to the lowest tested ratio of 1/10,000.

[0021] Figure 10 is a Python Code that may used to carry out methods described herein according to one embodiment.

DETAILED DESCRIPTION

[0022] Single molecule identifier adaptors and methods for their use are provided herein. According to the embodiments described herein, a single molecule identifier (SMI) adaptor molecule is provided. Said SMI adaptor molecule is double stranded, and may include a single molecule identifier (SMI) sequence, and an SMI ligation adaptor (Figure 2). Optionally, the SMI adaptor molecule further includes at least two PCR primer binding sites, at least two sequencing primer binding sites, or both.

[0023] The SMI adaptor molecule may form a "Y-shape" or a "hairpin shape." In some embodiments, the SMI adaptor molecule is a "Y-shaped" adaptor, which allows both strands to be independently amplified by a PCR method prior to sequencing because both the top and bottom strands have binding sites for PCR primers FC1 and FC2 as shown in the examples below. A schematic of a Y-shaped SMI adaptor molecule is also shown in Figure 2. A Y-shaped SMI adaptor requires successful amplification and recovery of both strands of the SMI adaptor molecule. In one embodiment, a modification that would simplify consistent recovery of both strands entails ligation of a Y-shaped SMI adaptor molecule to one end of a DNA duplex molecule, and ligation of a "U-shaped" linker to the other end of the molecule. PCR amplification of the hairpin-shaped product will then yield a linear fragment with flow cell sequences on either end. Distinct PCR primer binding sites (or flow cell sequences FC1 and FC2) will flank the DNA sequence corresponding to each of the two SMI adaptor molecule strands, and a given sequence seen in Read 1 will then have the sequence corresponding to the complementary DNA duplex strand seen in Read 2.

Mutations are scored only if they are seen on both ends of the molecule (corresponding to each strand of the original double-stranded fragment), i.e. at the same position in both Read 1 and Read 2. This design may be accomplished as described in the examples relating to double stranded SMI sequence tags.

[0024] In other embodiments, the SMI adaptor molecule is a "hairpin" shaped (or "U-shaped") adaptor. A hairpin DNA product can be used for error correction, as this product contains both of the two DNA strands. Such an approach allows for reduction of a given sequencing error rate N to a lower rate of N*N*(1/3), as independent sequencing errors would need to occur on both strands, and the same error among all three possible base substitutions would need to occur on both strands. For example, the error rate of 1/100 in the case of Illumina sequencing [32] would be reduced to (1/100)*(1/100)*(1/3) = 1/30,000.

[0025] An additional, more remarkable reduction in errors can be obtained by inclusion of a single-stranded SMI in either the hairpin adaptor or the "Y-shaped" adaptor will also function to label both of the two DNA strands. Amplification of hairpin-shaped DNA may be difficult as the polymerase must synthesize through a product containing significant regions of self-complementarity, however, amplification of hairpin-shaped structures has already been established in the technique of hairpin PCR, as described below. Amplification using hairpin PCR is further described in detail in U.S. Patent No. 7,452,699, the subject matter of which is hereby incorporated by reference as if fully set forth herein.

[0026] According to the embodiments described herein, the SMI sequence (or "tag") may be a double-stranded, complementary SMI sequence or a single-stranded SMI sequence. In some embodiments, the SMI adaptor molecule includes an SMI sequence (or "tag") of nucleotides that is degenerate or semi-degenerate. In some embodiments, the degenerate or semi-degenerate SMI sequence may be a random degenerate sequence. A double-stranded SMI sequence includes a first degenerate or semi-degenerate nucleotide n-mer sequence and a second n-mer sequence that is complementary to the first degenerate or semi-degenerate nucleotide n-mer sequence, while a single-stranded SMI sequence includes a first degenerate or semi-degenerate

nucleotide n-mer sequence. The first and/or second degenerate or semi-degenerate nucleotide n-mer sequences may be any suitable length to produce a sufficiently large number of unique tags to label a set of sheared DNA fragments from a segment of DNA. Each n-mer sequence may be between approximately 3 to 20 nucleotides in length. Therefore, each n-mer sequence may be approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length. In one embodiment, the SMI sequence is a random degenerate nucleotide n-mer sequence which is 12 nucleotides in length. A 12 nucleotide SMI n-mer sequence that is ligated to each end of a target nucleic acid molecule, as described in the Example below, results in generation of up to 4^{24} (i.e., 2.8 x 10^{14}) distinct tag sequences.

[0027] In some embodiments, the SMI tag nucleotide sequence may be completely random and degenerate, wherein each sequence position may be any nucleotide. (i.e., each position, represented by "X," is not limited, and may be an adenine (A), cytosine (C), guanine (G), thymine (T), or uracil (U)) or any other natural or non-natural DNA or RNA nucleotide or nucleotide-like substance or analog with base-pairing properties (e.g., xanthosine, inosine, hypoxanthine, xanthine, 7-methylguanine, 7-methylguanosine, 5,6-dihydrouracil, 5-methylcytosine, dihydouridine, isocytosine, isoguanine, deoxynucleosides, nucleosides, peptide nucleic acids, locked nucleic acids, glycol nucleic acids and threose nucleic acids). The term "nucleotide" as described herein, refers to any and all nucleotide or any suitable natural or non-natural DNA or RNA nucleotide or nucleotide-like substance or analog with base pairing properties as described above. In other embodiments, the sequences need not contain all possible bases at each position. The degenerate or semi-degenerate n-mer sequences may be generated by a polymerase-mediated method described in the Example below, or may be generated by preparing and annealing a library of individual oligonucleotides of known sequence. Alternatively, any degenerate or semi-degenerate n-mer sequences may be a randomly or non-randomly fragmented double stranded DNA molecule from any alternative source that differs from the target DNA source. In some embodiments, the alternative source is a genome or plasmid derived from bacteria, an organism other than that of the target DNA, or a combination of such alternative organisms or sources. The random or non-random fragmented DNA may be introduced into SMI adaptors to

serve as variable tags. This may be accomplished through enzymatic ligation or any other method known in the art.

[0028] In some embodiments, the SMI adaptor molecules are ligated to both ends of a target nucleic acid molecule, and then this complex is used according to the methods described below. In certain embodiments, it is not necessary to include n-mers on both adapter ends, however, it is more convenient because it means that one does not have to use two different types of adaptors and then select for ligated fragments that have one of each type rather than two of one type. The ability to determine which strand is which is still possible in the situation wherein only one of the two adaptors has a double-stranded SMI sequence.

[0029] In some embodiments, the SMI adaptor molecule may optionally include a double-stranded fixed reference sequence downstream of the n-mer sequences to help make ligation more uniform and help computationally filter out errors due to ligation problems with improperly synthesized adaptors. Each strand of the double-stranded fixed reference sequence may be 4 or 5 nucleotides in length sequence, however, the fixed reference sequence may be any suitable length including, but not limited to 3, 4, 5 or 6 nucleotides in length.

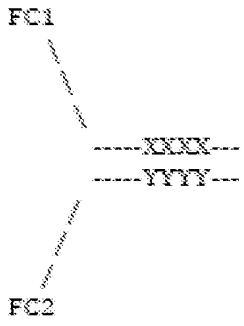
[0030] The SMI ligation adaptor may be any suitable ligation adaptor that is complementary to a ligation adaptor added to a double-stranded target nucleic acid sequence including, but not limited to a T-overhang, an A-overhang, a CG overhang, a blunt end, or any other ligatable sequence. In some embodiments, the SMI ligation adaptor may be made using a method for A-tailing or T-tailing with polymerase extension; creating an overhang with a different enzyme; using a restriction enzyme to create a single or multiple nucleotide overhang, or any other method known in the art.

[0031] According to the embodiments described herein, the SMI adaptor molecule may include at least two PCR primer or "flow cell" binding sites: a forward PCR primer binding site (or a "flow cell 1" (FC1) binding site); and a reverse PCR primer binding site (or a "flow cell 2" (FC2) binding site). The SMI adaptor molecule may also include at least two sequencing primer binding sites, each corresponding to a sequencing read. Alternatively, the sequencing primer binding sites may be added in a

separate step by inclusion of the necessary sequences as tails to the PCR primers, or by ligation of the needed sequences. Therefore, if a double-stranded target nucleic acid molecule has an SMI adaptor molecule ligated to each end, each sequenced strand will have two reads - a forward and a reverse read.

Double-stranded SMI sequences

[0032] Adaptor 1 (shown below) is a Y-shaped SMI adaptor as described above (the SMI sequence is shown as X's in the top strand (a 4-mer), with the complementary bottom strand sequence shown as Y's):

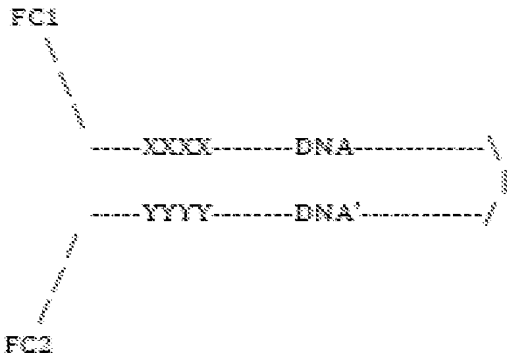


(Adaptor 1)

[0033] Adaptor 2 (shown below) is a hairpin (or "U-shaped") linker:

(Adaptor 2)

[0034] Following ligation of both adaptors to a double-stranded target nucleic acid, the following is structure is obtained:



[0035] When melted, the product will be of the following form (where "linker" is the sequence of adaptor 2):

FC1-------XXXX------DNA---------linker---------DNA'----------YYYY--------FC2

[0036] This product is then PCR amplified. The reads will yield:

Read 1:
XXXX-----DNA----

Read 2 (note that read 2 is seen as the complement of the bases sequenced:)
XXXX-----DNA----

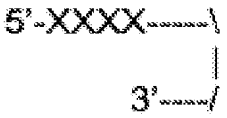
[0037] The sequences of the two duplex strands seen in the two sequence reads may then be compared, and sequence information and mutations will be scored only if the sequence at a given position matches in both of the reads.

[0038] This approach does not strictly require the use of an SMI tag, as the sheared ends can be used as identifiers to differentiate unique individual molecules from PCR duplicates. Thus the same concept would apply if one used any standard sequencing adaptor as "Adaptor 1" and the U---shaped linker as "Adaptor 2." However described below, there are a limited number of shear points flanking any given genomic position and thus the power to sequence deeply is increased via inclusion of the SMI tag. A hybrid method using a combination of sheared ends and a shorter n-mer tag (such as 1 or 2 or 3 or 4 or more degenerate or semi-degenerate bases) in the adaptor may also serve as unique molecular identifiers. Another design may include use of any sequencing adaptor (such as one lacking an n-mer tag) in conjunction with an n-mer tag that is incorporated into the U-shaped linker molecule. Such a design would be of the following form (where X and Y represent complementary degenerate or semi-degenerate nucleotides):

XXXX-----\
|
YYYY-----/

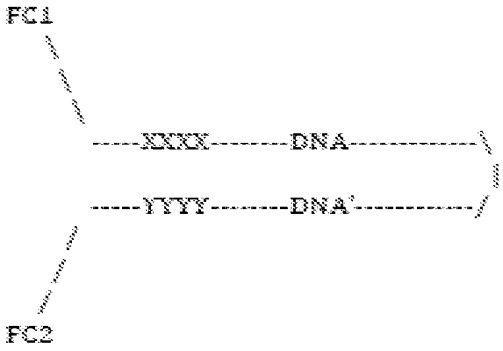
[0039] Synthesis of such a design may be obtained in a number of ways, for example synthesizing a set of hairpin oligonucleotides in which each individual oligonucleotide encodes a complementary n-mer sequence, or alternatively by using a

DNA polymerase to carry out extension from the following product (where X's represent degenerate nucleotides):

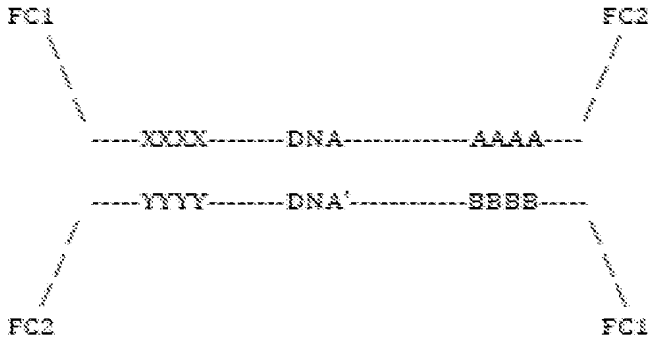


[0040] Inclusion of the SMI tag is also extremely useful for identifying correct ligation products, as the assay uses two distinct adaptors. This will yield multiple possible ligation products:

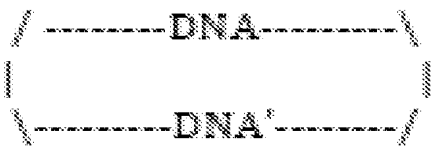
[0041] **Product I.** Adaptor 1---------DNA---------Adaptor 2, which yields the desired product:



[0042] **Product II.** Adaptor 1---------DNA---------Adaptor 1. This will result in the DNA being amplified as two separate strands, i.e. as occurs in the DCS approach described elsewhere in this document (the second copy of Adaptor 1 is shown below with the SMI as AAA – BBB to emphasize that every DCS adaptor has a distinct SMI sequence)



[0043] **Product III.** Adaptor 2---------DNA---------Adaptor 2. This will result in a non---amplifiable circular product shown below:



[0044] Product III is non-amplifiable, given the absence of primer binding sites and thus will not be present in the final DNA sequences. Thus only Product II needs to be avoided. The formation of Product II can be minimized in the ligation step by using an excess of Adaptor 2 (relative to Adaptor 1). Then primarily Products I and III will be obtained, with minimal formation of Product II. Additionally, a variety of biochemical means of enriching for products containing adaptor 2 are possible such as using affinity probes that are complementary to the hairpin loop sequence itself. Product I results in the same SMI sequence in both the Read 1 and Read 2 sequence reads. In the example depicted above, Product I sequences can thus be identified by virtue of having matching SMIs of the form XXXX in Read 1 and XXXX in Read 2.

[0045] By contrast, in the case of Product II, the SMI sequences on either end of the sequenced molecule will arise from distinct DCS adaptors having different SMI sequences. In the example shown above, Product II sequences yield SMIs of the form XXXX (Read 1) – BBB (Read 2) upon sequencing of the top strand, and BBBB (Read 1) – XXXX (Read 2) upon sequencing of the bottom strand. Thus Product II sequences can be easily identified and computationally removed from the final sequence data.

[0046] Data resulting from Product II is useful, because Product II corresponds to the product analyzed under the approach detailed in the Example below. Product I contains a self--complementary hairpin sequence that can impair polymerase extension during amplification, however, this type of amplification has already been enabled in the technique of “Hairpin PCR” [50] which involves linking of the two strands followed by amplification with gene-specific primers. Amplification conditions that are compatible with amplification of hairpin DNA are thus already established. Moreover, ligation and

amplification with circularizing "linkers" (i.e. hairpin linkers affixed to both ends of a fragment) has been demonstrated as a step in the Pacific Biosciences sample preparation workflow [49]. As the sequence of the linker itself does not matter in the workflow, the published linker sequences from either of these references would be adequate for use in the assay.

[0047] In some aspects of some embodiments, deliberate ligation of "U-shaped" adaptors or hairpin linkers containing 1) a double-stranded n-mer (or other form of degenerate or semi-degenerate double-stranded tag as enumerated above) plus 2) primer binding sites to both ends of a captured fragment may be desireable. Producing closed circles of captured material may help facilitate removal of non-captured DNA by exonuclease digestion given that circularized DNA will be protected from digestion by such enzymes. Additionally, closed circles may be pre-amplified using rolling circle amplification or serve as the substrate for continuous loop sequencing [49]. Recognition sites for restriction endonuclease digestion could be engineered into these adaptors to render closed loops open once again if more convenient for subsequent steps.

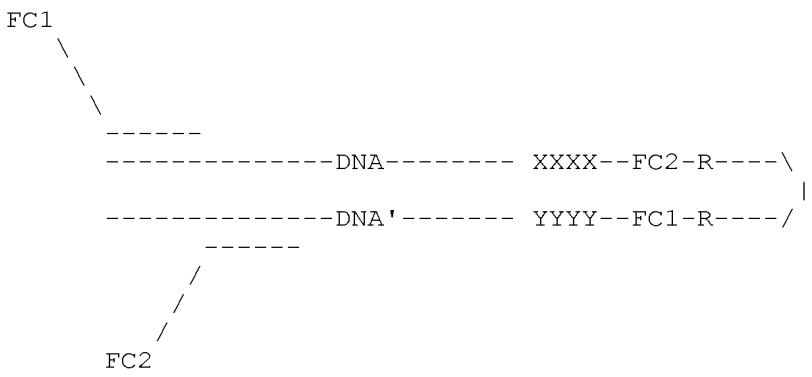
[0048] In another embodiment, flow cell sequences or PCR binding sites, again denoted as FC1 and FC2, may be included in both the PCR primers and the hairpin linker adaptor, as well as a ligatable sequence on the end of the hairpin linker (denoted as L below). The hairpin linker adaptor may additionally include one or more cleavable sequences, denoted as R in the example below (the R may be any appropriate restriction enzyme target sequence, or any other cleavable sequence). Such a hairpin linker design is shown below:

```
L-XXXX--FC2-R----\
                  |
L-YYYY--FC1-R----/
```

[0049] The target DNA with ligation site denoted as L is as follows:

```
--------------DNA--------L
--------------DNA'-------L
```

[0050] Following ligation of the linker, the product may be amplified with PCR primers as follows:



[0051] The resultant product will be of the form:

FC1--------DNA--------XXXX--FC2-R-----R-FC1--YYYY--------DNA'------- FC2

[0052] After amplification of the product, the cleavage sites R may be cleaved to result in the following sequencable products:

FC1--------DNA-------- XXXX--FC2

and

FC1--YYYY--------DNA'------- FC2

[0053] These products may then be sequenced directly. This design has the advantage of allowing for targeted sequencing of a specific region of the genome, and furthermore avoids the need to sequence a hairpin product, as sequencing of a hairpin will be less efficient due to the self-complementarity present within the hairpin molecule.

Single-stranded SMI sequences

[0054] In one embodiment, a single-stranded SMI sequence is incorporated into the single-stranded portion of the hairpin loop (regions of sequence complementarity are denoted as "="). The SMI sequence is shown as four nucleotides in length in the following examples, but in practice an Nmer of any length, including approximately 3 to 20 nucleotides, will suffice.

```
=====NNNN-----\
                |
=====----------/
```

[0055] Ligation of the hairpin linker and a Y-shaped sequencing adaptor (with PCR primer binding sites labeled as FC1 and FC2) yields the following product:

```
 FC1
    \
     \
      =====DNA =====NNNN------\
                              |
      =====DNA'=====----------/
     /
    /
   /
FC2
```

[0056] Melting and PCR amplification of this product yields the following DNA product:

```
FC1-------DNA-----NNNN------hairpin sequence------DNA'-------FC2
```

[0057] Following PCR duplication of the product and formation of consensus reads based upon the shared SMI sequence among all the PCR duplicates, the sequences of the two strands (denoted DNA and DNA') can then be compared to form a duplex consensus sequence.

[0058] In another embodiment, a single-stranded SMI is incorporated into a modified "Y-shaped" sequencing adaptor in which PCR primer binding sites are located at the sites labeled FC1 and FC2 (regions of sequence complementarity are depicted as "=")

```
FC1
   \
    \
     \--NNNN====DNA =======\
                           |
         FC2====DNA'=======/
```

[0059] It will be apparent to one skilled in the art that a single-stranded SMI sequence tag can be located in any of several positions within either the sequencing

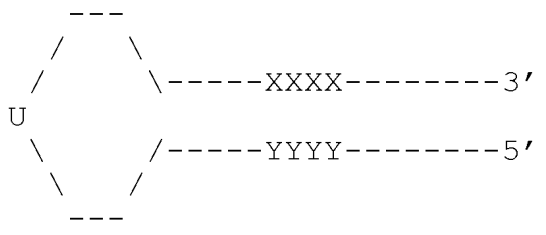
adaptor or the hairpin linker. The single-stranded SMI sequence tag can be synthesized as a random oligonucleotide sequence, or can be sequenced as a set of fixed sequences by synthesis on an array, or by any other suitable method known in the art.

Methods for synthesis of complementary or partially complementary double stranded SMI tags

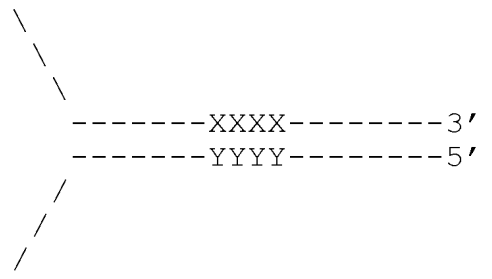
[0060] SMI adaptors molecules containing a double-stranded, complementary, degenerate or semi-degenerate SMI tag can be made by any of a number of methods, including copying of a single-stranded SMI sequence by a DNA polymerase as described above or synthesis and annealing of two oligonucleotides containing complementary SMI sequences. An additional method involves synthesizing a set of linear oligonucleotides which will self-anneal into the appropriate form. Inclusion of a cleavable linker in each oligonucleotide will then allow for conversion of a "hairpin shaped" SMI adaptor molecule into a "Y-shaped" SMI adaptor molecule. For example, an oligonucleotide may be prepared of the following form:

5'---------YYYY-------U-------XXXX----------3'

[0061] In this schematic, X and Y represent complementary nucleotides, and U indicates a cleavable linker, such as uracil (which can be cleaved by combined treatment with uracil DNA glycosylase and apurinic endonuclease), although any other cleavable linker will suffice. The oligonucleotide may be designed with appropriate regions of self-complementarity to anneal into the following form:



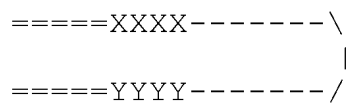
[0062] The linker (e.g. uracil) may then be cleaved, yielding a DCS adaptor:



[0063] A double-stranded SMI hairpin linker can be constructed by an analogous method but without the need for a cleavable linker. For example, a set of nucleotides of known sequence where X and Y represent the complementary SMI sequences can be synthesized on an array, or by any other suitable method known in the art:

5′=====XXXX-------------YYYY=====3′

[0064] This oligonucleotide can then self-anneal to form a hairpin linker with complementary SMI sequences.



[0065] Any of the oligonucleotides described above can also include any ligatable sequence as overhangs on either the 5' or 3' end, or can be used for blunt end ligation.

DCS SMI adaptor molecules may include sequences to allow for targeted DNA capture

[0066] DCS SMI adaptor molecules contain ligatable ends to allow attachment of the adaptor to a target DNA molecule. In some embodiments, the ligatable end may be complementary to a DNA overhang on the target DNA, for example, one generated by digestion of target DNA with a restriction endonuclease. Selective ligation of the adaptor to the targeted DNA containing the matching Single-stranded overhanging DNA sequence will then allow for partial purification of the targeted DNA. A non-limiting example of this embodiment is shown above in paragraphs [0048]-[0053]. In some

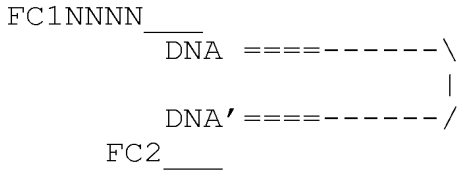
embodiments, the DCS SMI adaptor molecule, or a hairpin linker SMI adaptor molecule, may additionally contain modifications such as biotin to facilitate affinity purification of target DNA that has ligated to the adaptor.

[0067] In another embodiments, specific PCR primers can selectively amplify specific regions of genome when the adaptor that is ligated to the other end of the molecule is a hairpin (or "U-shape"). Alternatively, this method may be used with or without the need for this cleavable hairpin sequence.

Preparation of DNA for duplex consensus sequencing may be performed by PCR amplification in a hairpin structure

[0068] Another embodiment involves fragmentation of DNA at defined regions, for example by treatment of DNA with a site-specific restriction endonuclease or a mixture of such endonucleases, followed by annealing of a hairpin oligonucleotide linker, and amplification of the hairpin complex with PCR primers sufficient for amplification of the desired DNA sequence. Annealing of the hairpin linker to only one of the two ends of the DNA duplex could be accomplished by using different restriction enzymes to cut on either end of the target duplex, and then having the hairpin linker ligation adaptor being ligatable to only one of the two resultant ligatable ends.

[0069] The example shown below indicates forward and reverse PCR primers (labeled 1 and 2) in conjunction with a hairpin linker to allow linked amplification of both complementary strands of duplex DNA. Such amplification, in conjunction with a single-stranded or double-stranded SMI sequence, would allow for targeted amplification and high accuracy deep sequencing of a specific sequence of interest. In the schematic shown below, a single-stranded SMI sequence is incorporated into PCR primer FC1. It would be apparent to one skilled in the art that the SMI sequence could also be incorporated in primer FC2, or in the hairpin linker.



[0070] Amplified product:

FC1NNNN DNA----hairpin sequence----DNA'FC2

[0071] This product can then be subjected to consensus sequencing analysis. The SMI sequence allows one to group together products of PCR amplification arising from a single molecule of duplex DNA. The sequences of the two DNA strands can then be compared for error correction.

Uses of SMI adapter molecules

[0072] The SMI adaptor molecules described herein have several uses. In some embodiments, the SMI adaptor molecules described herein may be used in methods to obtain the sequence or other sequence-related information of a double-stranded target nucleic acid molecule. According to the embodiments described herein, the term "double-stranded target nucleic acid molecule" includes a double-stranded DNA molecule or a double-stranded RNA molecule. Thus, the SMI adaptor molecules and methods of use described herein are applicable to genotyping and other applications related to sequencing of DNA molecules, but are also applicable to RNA sequencing applications such as for sequencing of double-stranded RNA viruses. Methods for sequencing RNA may include any of the embodiments described herein with respect to DNA sequencing, and vice-versa. For example, any double stranded target nucleic acid molecule may be ligated to an SMI adaptor molecule which includes a double-stranded RNA or DNA n-mer tag and an RNA or DNA ligation adapter as described above. Methods exist for directly sequencing RNA [51]; alternatively, the ligated product may be reverse transcribed into DNA, and then sequenced as a double-stranded target DNA molecule.

[0073] In one embodiment, the double-stranded target nucleic acid molecule may be a sheared double-stranded DNA or RNA fragment. The sheared target DNA or RNA molecule may be end repaired and a double-stranded target nucleic acid sequence ligation adaptor may be added to each end of the sheared target DNA or RNA molecule. The double-stranded target nucleic acid sequence ligation adaptor may be any suitable ligation adaptor that is complementary to the SMI ligation adaptor described above including, but not limited to a T-overhang, an A-overhang, a CG overhang, blunt end or

any other ligatable sequence. In some embodiments, the double-stranded target nucleic acid sequence ligation adaptor may be made using a method for A-tailing or T-tailing with polymerase extension; adding an overhang with a different enzyme; using a restriction enzyme to create a ligatable overhang; or any other method known in the art.

[0074] Methods to obtain the sequence or other sequence-related information of a double-stranded target nucleic acid molecule may include a step of ligating the double-stranded target nucleic acid molecule to at least one SMI adaptor molecule, such as those described above, to form a double-stranded target nucleic acid complex. In one embodiment, each end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule. The double-stranded target nucleic acid complex is then amplified by a method known in the art (e.g., a PCR or non-PCR method known in the art), resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products. These products are then sequenced using any suitable method known in the art including, but not limited to, the Illumina sequencing platform, ABI SOliD sequencing platform, Pacific Biosciences sequencing platform, 454 Life Sciences sequencing platform, Ion Torrent sequencing platform, Helicos sequencing platform, and nanopore sequencing technology.

[0075] In certain embodiments, a method of generating an error corrected double-stranded consensus sequence is provided. Such a method, also referred to as duplex consensus sequencing (DCS), allows for a quantitative detection of sites of DNA damage. DCS analysis facilitates the detection of DNA damage signatures, in that single stranded DNA mutations that are not present in the complementary strand can be inferred to be artifactual mutations arising from damaged nucleotides. Not only can one correct for these erroneous mutations, but the ability to indirectly infer that damage is present on the DNA could be a useful biomarker (e.g. for cancer risk, cancer metabolic state, mutator phenotype related to defective damage repair, carcinogen exposure, chronic inflammation exposure, individual-specific aging, neurodegenerative diseases etc). The ability to use different polymerases during the first round(s) of PCR to mis-incorporate at damage sites could potentially add even more information. Besides polymerases, other DNA modifying/repair enzymes could be used prior to amplification to convert damage of one sort that doesn't give a specific mutagenic signature into

another sort that does with whatever polymerase is used. Alternatively, DNA modifying/repair enzymes could be used to remove damaged bases, and one could sequence both strands of DNA both with and without the enzymatic treatment. Mutations in single-stranded DNA that are seen to be removed by the enzymatic treatment can thus be inferred to be arising due to DNA damage. This could be useful on human nuclear or mtDNA but also might also be useful with model organisms (mice, yeast, bacteria etc), treated with different new damaging agents, facilitating a screen for DNA damaging compounds that would be analogous to the widely used Ames test [52].

[0076] The method of generating an error corrected double-stranded consensus sequence may include a first stage termed "single strand consensus sequencing" (SSCS) followed by a second stage of duplex consensus sequencing (DCS). Therefore, the method includes steps of tagging individual duplex DNA molecules with an SMI adaptor molecule, such as those described above; generating a set of PCR duplicates of the tagged DNA molecules by performing a suitable PCR method; creating a single strand consensus sequence from all of the PCR duplicates which arose from an individual molecule of single-stranded DNA. Each DNA duplex should result in two single strand consensus sequences. The work through these three steps conclude the first stage and is termed SSCS.

[0077] The method of generating an error corrected double-stranded consensus sequence further comprises the second stance that is termed DCS. The DCS stage includes steps of comparing the sequence of the two single strand consensus sequences arising from a single duplex DNA molecule, and further reducing sequencing or PCR errors by considering only sites at which the sequences of both single-stranded DNA molecules are in agreement. The method that includes the first stage and the second stage termed Duplex Consensus Sequencing (DCS).

[0078] The step of tagging of both strands of individual duplex DNA may be accomplished by ligation of degenerate or semi-degenerate complementary DNA sequences; as the complementary nature of the two strands of such a tag sequence allows the two molecules to be grouped together for error correction. Alternatively, as described above, the two duplex DNA strands may be linked by ligation of a U-shaped

SMI adaptor molecule, and the two DNA strands can thus both be tagged with a single-stranded SMI tag.

[0079] In the method described above, a set of sequenced SMI-DNA products generated in the methods described above may be grouped into families of paired target nucleic acid strands based on a common set of SMI sequences. Then, the paired target nucleic acid strands can be filtered to remove nucleotide positions where the sequences seen on both of the paired partner DNA strands are not complementary. This error corrected double-stranded consensus sequence may be used in a method for confirming the presence of a true mutation (as opposed to a PCR error or other artifactual mutation) in a target nucleic acid sequence. According to certain embodiments, such a method may include identifying one or more mutations present in the paired target nucleic acid strands that have one or more nucleotide positions that disagree between the two strands, then comparing the mutation present in the paired target nucleic acid strands to the error corrected double-stranded consensus sequence. The presence of a true mutation is confirmed when the mutation is present on both of the target nucleic acid strands and also appear in all members of a pared target nucleic acid family.

[0080] The accuracy of current approaches to next-generation sequencing is limited due to their dependence on interrogating single-stranded DNA. This dependence makes potential sources of error such as PCR amplification errors and DNA damage fundamentally limiting. However, the complementary strands of a double-stranded DNA molecule (or "DNA duplex") contain redundant sequencing information (i.e., one molecule reciprocally encoding the sequence information of its partner) which can be utilized to eliminate such artifacts. Limitations related to sequencing single-stranded DNA (e.g., sequencing errors) may therefore be overcome using the methods described herein. This is accomplished by individually tagging and sequencing each of the two strands of a double-stranded (or duplex) target nucleic acid molecule and comparing the individual tagged amplicons derived from one half of a double-stranded complex with those of the other half of the same molecule. Duplex Consensus Sequencing (DCS), significantly lowers the error rate of sequencing. In some embodiments, the DCS method may be used in methods for high sensitivity detection of

rare mutant and variant DNA as described further below.

[0081] As described above, one approach that has previously been reported for DNA sequencing involves incorporation of a random tag sequence into a PCR primer [36]. This approach results in an improvement in accuracy relative to standard Illumina sequencing, but is fundamentally limited in that it is based upon amplification and sequencing of single-stranded DNA and thus cannot overcome limitations in sensitivity owing to single-stranded DNA damage events. In the methods described herein, PCR duplicates are generated from a single strand of DNA, and the sequences of the duplicates are compared. Mutations are scored only when they are present in multiple replicates of a single starting molecule. The DCS approach overcomes the limitation of previous approaches by considering both DNA strands.

[0082] DNA damage should not be a limiting factor in DCS, because miscoding damage events at a single base-pair position occur essentially exclusively on only one of the two DNA strands. For DNA damage to result in an artifactual mutation in DCS, damage would need to be present at the same nucleotide position on both strands. Even if complementary nucleotides in a duplex were both damaged, the damage would need to result in complementary sequencing errors to result in mis-scoring of a mutation. Likewise, spontaneous PCR errors would need to result in complementary mutations at the same position on both strands; with a first-round mutation frequency of Taq polymerase of approximately 10^{-5} and three possible incorrect bases that could be mis-inserted, the probability of two complementary PCR errors occurring would be $10^{-5} \times 10^{-5} \times 1/3 = 3.3 \times 10^{-11}$

[0083] According to some embodiments, the sequencing method may be performed using the Illumina or similar platforms including those enumerated above without the use of SMI adaptor molecules, but instead by using the random shear points of DNA as identifiers. For a given DNA sequence seen in sequencing read 1 with a specific set of shear points, the partner strand will be seen as a matching sequence in read two with identical shear points. In practice, this approach is limited by the limited number of possible shear points that overlap any given DNA position. However, according to some embodiments, shear points of a target nucleic acid molecule may be

used as unique identifiers to identify double-stranded (or duplex) pairs, resulting in an apparent error frequency at least as low as that seen with traditional sequencing methods, but with a significantly lower loss of sequence capacity. In other embodiments, DCS based on shear points alone may have a role for confirmation that specific mutations of interest are true mutations which were indeed present in the starting sample (i.e. present in both DNA strands), as opposed to being PCR or sequencing artifacts. Overall, however, DCS is most generally applicable when randomized, complementary double-stranded SMI sequences are used. A 24 nucleotide double-stranded SMI sequence was used in the Example described below, which may yield up to 4^{24} = 2.8 x 10^{14} distinct double-stranded SMI sequences. Combining information regarding the shear points of DNA with the SMI tag sequence would allow a shorter SMI to be used, thus minimizing loss of sequencing capacity due to sequencing of the SMI itself.

[0084] In certain embodiments, the SMI adaptor molecules may also be used in methods of single-molecule counting for accurate determination of DNA or RNA copy number [38]. Again, since the SMI tags are present in the adaptors, there are no altered steps required in library preparation, which is in contrast to other methods for using random tags for single-molecule counting. Single-molecule counting has a large number of applications including, but not limited to, accurate detection of altered genomic copy number (e.g., for sensitive diagnosis of genetic conditions such as trisomy 21 [47]), for accurate identification of altered mRNA copy number in transcriptional sequencing and chromatin immunoprecipitation experiments, quantification of circulating microRNAs, quantification of viral load of DNA or RNA viruses, quantification of microorganism abundance, quantification of circulating neoplastic cells, counting of DNA-labeled molecules of any variety including tagged antibodies or aptamers, and quantification of relative abundances of different individual's genomes in forensic applications.

[0085] In another embodiment, the SMI adaptor molecules may be used in methods for unambiguous identification of PCR duplicates. In order to restrict sequencing analysis to uniquely sequenced DNA fragments, many sequencing studies include a step to filter out PCR duplicates by using the shear points at the ends of DNA

molecules to identify distinct molecules. When multiple molecules exhibit identical shear points, all but one of the molecules are discarded from analysis under the assumption that the molecules represent multiple PCR copies of the same starting molecule. However sequence reads with identical shear points can also reflect distinct molecules because there are a limited number of possible shear points at any given genomic location, and with increasing sequencing depth, recurrent shear points are increasingly likely to be seen [48]. Because the use of SMI tags (or "double-stranded SMI sequences") allows every molecule to be uniquely labeled prior to PCR duplication, true PCR duplicates may be unambiguously identified by virtue of having a common (i.e., the same or identical) SMI sequence. This approach would thereby minimize the loss of data by overcoming the intrinsic limitations of using shear points to identify PCR duplicates.

[0086] Importantly, once SMI-containing adaptors are synthesized by a straightforward series of enzymatic steps or are produced through synthesis of a set of oligonucleotides containing complementary tag sequences, they may be substituted for standard sequencing adaptors. Thus, use of DCS does not require any significant deviations from the normal workflow of sample preparation for Illumina DNA sequencing. Moreover, the DCS approach can be generalized to nearly any sequencing platform because a double-stranded SMI tag can be incorporated into other existing adaptors, or for sequencing approaches that do not require adaptors, a double-stranded SMI tag can be ligated onto duplex DNA sample prior to sequencing. The compatibility of DCS with existing sequencing workflows, the potential for greatly reducing the error rate of DNA sequencing, and the multitude of applications for the double-stranded SMI sequences validate DCS as a technique that may play a general role in next generation DNA sequencing.

[0087] The following examples are intended to illustrate various embodiments of the invention. As such, the specific embodiments discussed are not to be construed as limitations on the scope of the invention. It will be apparent to one skilled in the art that various equivalents, changes, and modifications may be made without departing from the scope of invention, and it is understood that such equivalent embodiments are to be included herein. Further, all references cited in the disclosure are hereby incorporated

by reference in their entirety, as if fully set forth herein.

EXAMPLES

Example 1: Generation of SMI Adaptor Molecules and their use in sequencing double-stranded target DNA

Materials and Methods

[0088] *Materials.* Oligonucleotides were from IDT and were ordered as PAGE purified. Klenow exo- was from NEB. T4 ligase was from Enzymatics.

[0089] *DNA isolation.* Genomic DNA was isolated from normal human colonic mucosa by sodium iodide extraction (Wako Chemicals USA).

[0090] *Adaptor synthesis.* The adaptors were synthesized from two oligos, designated as:

the primer strand:
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT (SEQ ID NO:1); and
the template strand:
/5phos/ACTGNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTG AACTCCAGTCAC (SEQ ID NO:2).

[0091] The two adaptor strands were annealed by combining equimolar amounts of each oligo to a final concentration of 50 micromolar and heating to 95°C for 5 minutes. The oligo mix was allowed to cool to room temperature for over 1 hour. The annealed primer-template complex was extended in a reaction consisting of 40 micromolar primer-template, 25 units Klenow exo-DNA polymerase (New England Biolabs), 250 micromolar each dNTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, and 1 mM dithiothreitol (DTT) for 1 hour at 37°C. The product was isolated by ethanol precipitation. Due to the partial A-tailing property of Klenow exo-, this protocol results in a mixture of blunt-ended adapters and adapters with a single-nucleotide A hverhang. A single-nucleotide A overhang was added to residual blunt fragments by incubating the adapters with 25 units Klenow exo-, 1 mM dATP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM MgCl2, and 1 mM dithiothreitol (DTT) for 1 hour at 37°C. The product was again ethanol precipitated and resuspended to a final concentration of 50 micromolar.

[0092] *Sequencing library preparation.* 3 micrograms of DNA was diluted into

130 microliters of TE buffer (10 mM tris-HCl, pH 8.0, 0.1 M EDTA) and was sheared on the Covaris AFA system with duty cycle 10%, intensity 5, cycles/burst 200, time 20 seconds x 6, temperature 4°C. DNA was purified with 2 volumes of Agencourt AMPure XP beads per the manufacturer's protocol. After end-repair with the NEB end-repair kit per the manufacturer's protocol, DNA fragments larger than the optimal range of ~200-500 bp were removed by adding 0.7 volumes of AMPure XP beads and transferring the supernatant to a separate tube (fragments larger than 500 bp bind to the beads and are discarded). An additional 0.65 volumes of AMPure XP beads were added (this step allows fragments of approximately 200bp or greater to bind to the beads). The beads were washed and DNA eluted. DNA was then T-tailed in a reaction containing 5 units Klenow exo-, 1 mM dTTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM MgCl2, 1 mM. The reaction proceeded for 1 hour at 37 C. DNA was purified with 1.2 volumes of AMPure XP beads. The custom adaptors were ligated by combining 750 ng of T-tailed DNA with 250 pmol adaptors in a reaction containing 3000 units T4 DNA ligase, 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 5 mM DTT, 1 mM ATP. The reaction was incubated 25 C for 15 minutes, and purified with 1.2 volumes of AMPure XP beads.

[0093] *Pre-capture amplification.* 375 ng adaptor-ligated DNA was PCR amplified with primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and GTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:4) using the Kappa high-fidelity PCR kit for 8 cycles with an annealing temperature of 60 C. The product was purified with 1.2 volumes of AMPure XP beads.

[0094] *DNA capture.* Target capture was performed with the Agilent SureSelect system per the manufacturer's recommendations, except that capture volumes were performed at one-half of the standard volume. The capture set targeted an arbitrary 758 kb region of the genome consisting of both coding and noncoding sequences. Capture baits were 120 nt in length, and were prepared with the Agilent eArray tool with 3x tiling.

[0095] *Post-capture amplification.* Captured DNA was amplified with PCR primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and CAAGC AGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGC (SEQ ID

NO:5) where XXXXXX indicates the position of a fixed multiplexing barcode sequence). 2 0 fmol of DNA was used per lane for sequencing on an Illumina HiSeq 2000.

[0096] *Data processing.* Reads with intact SMI adaptors include a 12 nucleotide random sequence, followed by a 5 nucleotide fixed sequence. These reads were identified by filtering out reads that lack the expected fixed sequence at positions 13-17. The SMI sequence from both the forward and reverse sequencing reads (i.e., the first and second degenerate n-mer sequences) was computationally added to the read header, and the fixed sequence removed. The first 4 nucleotides located following the adaptor sequence were also removed due to the propensity for ligation and end-repair errors to result in an elevated error rate near the end of the DNA fragments. Reads having common (i.e., identical) SMI sequences were grouped together, and were collapsed to generate a consensus read. Sequencing positions were discounted if the consensus group covering that position consisted of fewer than 3 members, or if fewer than 90% of the sequences at that position in the consensus group had the identical sequence. Reads were aligned to the human genome with the Burrows-Wheeler Aligner (BWA). The consensus sequences were then paired with their strand-mate by grouping each 24 nucleotide tag of form AB in read 1 with its corresponding tag of form BA in read 2. Resultant sequence positions were considered only when information from both DNA strands was in perfect agreement. An overview of the data processing workflow is as follows:

1. Discard reads that do not have the 5 nt fixed reference (or "spacer") sequence (CAGTA; SEQ ID NO:6) present after 12 random nucleotides.
2. Combine the 12 nt SMI tags from read 1 and read 2, and transfer the combined 24 nt SMI sequence into the read header.
3. Discard SMIs with inadequate complexity (i.e., those with > 10 consecutive identical nucleotides).
4. Remove the 5 nt fixed reference sequence.
5. Trim an additional 4 nt from the 5' ends of each read pair (sites of error prone end repair).
6. Group together reads which have identical 24 nt SMIs.
7. Collapse to SMI consensus reads, scoring only positions with 3 or more SMI duplicates and >90% sequence identity among the duplicates.
8. For each read in read 1 file having SMI of format AB, group with corresponding DCS partner in read 2 with SMI of format BA.
9. Only score positions with identical sequence among both DCS partners.
10. Align reads to the human genome.

[0097] Code for carrying out the workflow may be pre-existing or may involve programming within the skill of those in the art. In some embodiments, however, the Python code, which is illustrated in Figure 10, may be used for carrying out the pairing and scoring of partner strands according to steps 8 and 9 of the workflow described above.

Overview

[0098] To overcome limitations in the sensitivity of variant detection by single-stranded next-generation DNA sequencing, an alternative approach to library preparation and analysis was designed, which is known herein as Duplex Consensus Sequencing (DCS) (Figure 1). The DCS method described herein involves tagging both strands of duplex DNA with a random, yet complementary double-stranded nucleotide sequence, which is known herein as a double-stranded single molecule identifier (SMI) sequence. The SMI sequences (in this case, double stranded SMI sequences) are incorporated into the SMI adaptor molecules by introducing a single-stranded randomized nucleotide sequence into one adapter strand and the extending the opposite strand with a DNA polysmerase to yield a complementary, double-stranded SMI sequence (Figure 2). The individually tagged strands are then PCR amplified. Every duplicate that arises from a single strand of DNA will have the same SMI, and thus each strand in a DNA duplex pair generates a distinct, yet related population of PCR duplicates after amplification owing to the complementary nature of the SMIs on the two strands of the duplex. Comparing the sequence obtained from each of the two strands comprising a single molecule of duplex DNA facilitates differentiation of sequencing errors from true mutations. When an apparent mutation is, due to a PCR or sequencing error, the substitution will only be seen on a single strand. In contrast, with a true DNA mutation, complementary substitutions will be present on both strands (see Figure 4C).

[0099] Following tagging with a double-stranded SMI and PCR amplification, a family of molecules is obtained that arose from a single DNA molecule; members of the same PCR "family" are then grouped together by virtue of having a common (i.e., the same) SMI tag sequence. The sequences of uniquely tagged PCR duplicates are

subsequently compared in order to create a PCR consensus sequence. Only DNA positions that yield the same DNA sequence in a specified proportion of the PCR duplicates in a family, such as 90% of the duplicates in one embodiment, are used to create the PCR consensus sequence. This step filters out random errors introduced during sequencing or PCR to yield the PCR consensus sequences, each of which derives from an individual molecule of single-stranded DNA. This set of PCR consensus sequences are called single strand consensus sequences (SSCSs).

[00100] Next, PCR consensus sequences arising from two complementary strands of duplex DNA can be identified by virtue of the complementary SMIs (Figure 3) to identify the "partner SMI." Specifically, a 24-nucleotide SMI consists of two 12-nucleotide sequences that can be designated XY. For an SMI of form XY in read 1, the partner SMI will be of form YX in read 2. An example to illustrate this point is given in Figure 4. Following partnering of two strands by virtue of their complementary SMIs, the sequences of the strands are compared. Sequence reads at a given position are kept only if the read data from each of the two paired strands is in agreement.

Results

[00101] In order to label or tag each of the strands of duplex DNA with unique complementary tags, adaptors which contain the standard sequences required for the Illumina HiSeq system were synthesized, but with addition of a double-stranded, complementary SMI sequence (or "tag") of 12 random nucleotides (or a random "degenerate sequence") per strand. Target DNA molecules having a random SMI sequence n-mer that is 12 nucleotides in length on each end will therefore have a unique 24 nucleotide SMI sequence. The adaptors were prepared (Figure 2) from two partially complementary oligonucleotides, one of which has a single-stranded 12 nucleotide random nucleotide sequence (i.e. a first random degenerate nucleotide n-mer sequence) followed by a single stranded fixed reference sequence that is 4 nucleotides in length. The single-stranded random nucleotide tag was converted to a double-stranded, complementary SMI tag by extension with Klenow exo- DNA polymerase and the extended adaptor was purified by ethanol precipitation. Due to the partial A-tailing property of Klenow exo-, this protocol results in a mixture of blunt-ended adaptors and adaptors with a single-nucleotide A overhang (data not shown). A single-

nucleotide A-overhang was added to the residual blunt fragments by incubating the adaptors with Klenow exo- DNA polymerase and a high concentration of dATP (1 mM), and purified the adaptors again by ethanol precipitation.

[00102] DNA for sequencing was sheared and end-repaired by standard methods, with size-selection for fragments in the range of ~200-500 bp by size-selective binding to Ampure XP beads. Standard Illumina library preparation protocols involve ligating A-tailed DNA to T-tailed adaptors. However, because A-tailed adaptors were used, the DNA was T-tailed by incubating the end-repaired DNA with Klenow exo- DNA polymerase and 1 mM dTTP. The adaptor-ligated library was PCR amplified and subjected to SureSelect capture, with targeting of an arbitrary 758 kb portion of the genome (DNA coordinates available upon request). The efficiency of adaptor ligation, PCR amplification, DNA capture, and sequencing were comparable to those seen with standard library preparation methods (data not shown). Although Agilent Sure Select probes are used in this example, any suitable method of DNA selection may be used to capture particular target double-stranded DNA sequences. For example, selection and capture may be accomplished by any selection by hybridization method (e.g., Agilent SureSelect, Primer Extension Capture, exploitation of biotinylated PCR amplicons as bait, Agilent HaloPlex) wherein probes that target the desired double-stranded DNA sequence may be recovered by an in-array capture (using probes immobilized on glass slides) or by affinity using magnetic beads in an in-solution capture. In addition, mitochondrial and some other forms of DNA may be isolated by size selection. Alternatively, in some embodiments, no enrichment is performed.

[00103] This protocol was used to sequence DNA isolated from normal colonic mucosa. Mutations were initially scored without consideration of the SMI sequences. PCR duplicates were filtered out with samtools rmdup, a standard tool which uses the shear points of DNA molecules to identify PCR duplicates, as molecules arising from duplicated DNA will have shared shear points. In order to focus specifically on non-clonal mutations, only those positions in the genome with at least 20x coverage and at which fewer than 5% of reads differed from the hg19 reference sequence were considered. This approach resulted in 70.9 million nucleotides of sequence data and 56,890 mutations, indicating an overall mutation frequency of 8.03×10^{-4}, in accord with

the error rate of Illumina next-generation sequencing of ~0.1-1% [32].

[00104] Next, the SMI tags were used to group together PCR duplicates that arose from individual single-stranded DNA molecules and to create a consensus sequence from the family of duplicates. At least 3 PCR duplicates were required, with at least 90% agreement in sequence among all duplicates, to consider a site for mutations. Scoring the mutation frequency as above, again considering only sites with a minimum of 20x coverage and with <5% of reads differing from reference, resulted in 145 million nucleotides of sequence with 6,508 mutations and an overall mutation frequency of 4.47×10^{-5}, consistent with prior reports [36]. Notably, far more nucleotides of DNA sequence were obtained in this approach (145 million) than in the standard Illumina sequencing approach (70 million) detailed above which is dependent on use of the shear points of single-ended reads to identify PCR duplicates. The improved sequence coverage arose from use of the SMI to identify PCR duplicates, because identifying PCR duplicates by consideration of uniquely sheared DNA ends is fundamentally limited by the small number of possible shear points that overlap a given position of the genome and the propensity for specific genomic regions to be more readily undergo shearing. Thus filtering PCR duplicates by using shear points resulted in discarding a large portion of the reads.

[00105] Finally, the complementary nature of the double-stranded SMI sequences was used to identify pairs of consensus groups that arose from complementary DNA strands. Sequence reads were considered only when the read data from each of the two strands is in perfect agreement. In a pilot experiment, after grouping of PCR duplicates as above, 29,409 SMI partner pairs were found, indicative that fewer than 1% of tags had their corresponding partner tag present in the library. The low recovery of tag pairs was most likely due to inadequate amplification of the starting DNA library. Among these tag-pairs, 24,772 duplex consensus strands were identified with an average strand length of 82 nucleotides, resulting in 2 million nucleotides of DNA consensus sequence. The sequences of the paired duplex strands disagreed at 3,585 of the nucleotide positions, indicative of single-stranded errors (i.e. PCR or sequencing errors); these sites of disagreement were removed, leaving only bases at which the sequence of both duplex strands were in perfect agreement. Next, as above, analysis

of mutation frequencies was restricted to sites with at least 10x coverage and at which fewer than 10% of reads disagreed from the hg19 reference sequence. Because the 2 million nucleotides of read data were spread across a 758kb target, our average depth was only ~3x. Thus only 14,464 nucleotides of DNA sequence corresponded to sites with at least 10x depth. Among these sites, zero mutations were seen. To increase the number of tag pairs considered, analysis described above was repeated, but PCR duplicates were grouped with a minimum of only 1 duplicate per site. This resulted in 28,359 nucleotides of DNA sequence with at least 10x depth. Again, no mutations were detected.

[00106] Current experiments are being performed on vastly smaller target DNA molecules (ranging from ~300 bp to ~20 kb in size). Use of smaller DNA targets will allow for much greater sequencing depth, and far more accurate assessment of the background mutation rate of the assay. In addition, the protocol has been modified to incorporate a greater number of PCR cycles initiated off a smaller number of genome equivalents, which will increase the fraction of tags for which both of the partner tag strands have been sufficiently amplified to be represented in the final sequence data. Indeed, among the 3.6 million SMIs present in our initial library which underwent PCR duplication, 1.5 million of the SMIs were present only once, indicating insufficient amplification of the DNA due in part to the low number of PCR cycles used.

Example 2: Duplex Sequencing of Human Mitochondrial DNA.

Materials and methods

[00107] In addition to those described in Example 1 above, the following materials and methods were also used.

[00108] *DNA Isolation*. Mitochondrial DNA was isolated as previously described (4).

[00109] *Data processing*. The entire human genome sequence (hg19) was used as reference for the mitochondrial DNA experiment, and reads that mapped to chromosomal DNA were removed. Reads sharing identical tag sequences were then grouped together and collapsed to consensus reads. Sequencing positions were

discounted if the consensus group covering that position consisted of fewer than three members or if fewer than 90% of the sequences at that position in the consensus group had the identical sequence. A minimum group size of three was selected because next-generation sequencing systems have an average base calling error rate of ~1/100. Requiring the same base to be identified in three distinct reads decreases the frequency of single-strand consensus sequence (SSCS) errors arising from base-call errors to $(1/100)3 = 1 \times 10^{-6}$, which is below the frequency of spontaneous PCR errors that fundamentally limit the sensitivity of SSCSs. The requirement for 90% of sequences to agree to score a position is a highly conservative cutoff. For example, with a group size of eight, a single disagreeing read will lead to 87.5% agreement and the position will not be scored. If all groups in an experiment are of size nine or less, this cutoff will thus require perfect agreement at any given position to score the position. Further development of our protocol may allow for less stringent parameters to be used to maximize the number of SSCS and duplex consensus sequence (DCS) reads that can be obtained from a given experiment.

Results

[00110] Having established the methodology for Duplex Sequencing with M13mp2 DNA, which is a substrate for which the mutation frequency and spectrum are fairly well established, it was desired to apply the approach to a human DNA sample. Thus, mitochondrial DNA was isolated from human brain tissue and sequenced the DNA after ligation of Duplex Sequencing adapters. A standard sequencing approach with quality filtering for a Phred score of 30 resulted in a mutation frequency of 2.7 × 10−3, and SSCS analysis yielded a mutation frequency of 1.5 × 10−4. In contrast, DCS analysis revealed a much lower overall mutation frequency of 3.5 × 10−5 (Figure 5A). The frequency of mutations in mitochondrial DNA has previously been difficult to measure directly due in part to sources of error in existing assays that can result in either overestimation or underestimation of the true value. An additional confounder has been that most approaches are limited to interrogation of mutations within a small fraction of the genome [56]. The method of single-molecule PCR, which has been proposed as an accurate method of measuring mitochondrial mutation frequency [56] and is considered resistant to damage-induced background errors [57], has resulted in a reported

mitochondrial mutation frequency in human colonic mucosa of 5.9 × 10−5 ± 3.2 × 10−5 [56], which is in excellent agreement with our result. Likewise, mitochondrial DNA sequence divergence rates in human pedigrees are consistent with a mitochondrial mutation frequency of 3–5 × 10−5 [58, 59].

[00111] When the distribution of mutations throughout the mitochondrial genome is considered, the quality filtered reads (analyzed without consideration of the tags) have many artifactual errors, such that identification of mutational hotspots is difficult or impossible (Figure 5B). DCS analysis removed these artifacts (Figure 5C) and revealed striking hypermutability of the region of replication initiation (D loop), which is consistent with prior estimates of mutational patterns in mitochondrial DNA based upon sequence variation at this region within the population [60].

[00112] SSCS analysis produced a strong mutational bias, with a 130-fold excess of G→T relative to C→A mutations (Figure 5D), consistent with oxidative damage of the DNA leading to first-round PCR mutations as a significant source of background error. A high level of oxidative damage is expected in mitochondrial DNA, due to extensive exposure of mitochondria to free radical species generated as a byproduct of metabolism [61]. DCS analysis (Figure 5E) removed the mutational bias and revealed that transition mutations are the predominant replication errors in mitochondrial DNA. The DCS mutation spectrum is in accord with prior estimates of deamination events [62] and T-dGTP mispairing by the mitochondrial DNA polymerase [63] as primary mutational forces in mitochondrial DNA. Furthermore, the mutation spectrum of our mitochondrial data are consistent with previous reports of heteroplasmic mutations in human brain showing an increased load of A→G/T→C and G→A/C→T transitions, relative to transversions [64, 65]. A similar spectral bias has also been reported in mice [62, 66] and in population studies of Drosophila melanogaster [67].

Example 3: Demonstration of error-correction by DCS using randomly sheared DNA ends as Single Molecule Identifiers

Materials and Methods

[00113] In addition to those described in the Examples above, the following materials and methods were also used to demonstrate the capability of DCS analysis to

remove sequencing errors

[00114] *Sequencing library preparation.* Genomic DNA was isolated from a derivative of *Saccharomyces cerevisiae* strain SC288 by standard methods. The DNA was randomly sheared by the Covaris AFA system, followed by end-repair, A-tailing, and ligation of Illumina TruSeq DNA sequencing adaptors, all by standard library preparation methods. The resultant sequence data consisted of an average 32.5 fold depth of the 12 megabase *S. cerevisiae* genome.

[00115] *Data analysis.* The first 10 nucleotides of each sequencing read pair, corresponding to the randomly sheared DNA ends, were combined, such that the first 10 nucleotides of read 1, referred to as A, was combined with the first 10 nucleotides of read 2, referred to as B, to yield an SMI tag of form AB. Reads were grouped according to SMI sequence, and nucleotide reads were considered only if they agreed among at least 90% of family members sharing a given tag sequence. For DCS analysis, a tag of form AB1 is partnered with the corresponding tag of form BA2, and nucleotide positions are considered only when the sequence is in agreement among read pairs with matching tags AB1 and BA2.

Results

[00116] In order to demonstrate the capability of DCS analysis to remove sequencing errors, a sequencing library was prepared under standard conditions with commercially available sequencing adaptors, and the randomly sheared DNA ends were used as SMI's. First, reads were grouped by SMI with a minimum family size of 1 member. Considering only sites with a minimum of 20x coverage and with <5% of reads differing from reference, this analysis resulted in 644.8 million nucleotides of sequence data and 2,381,428 mutations, yielding an overall mutation frequency of 3.69 x 10^{-3}.

[00117] The data was then subjected to DCS analysis with the SMI tags, searching for tags of form AB1 that have partner tags of form BA2, and considered only positions at which the sequence from the two strands was in perfect agreement. 3.1% of the tags had a matching partner present in the library, resulting in 2.9 million nucleotides of sequence data. The sequences of the duplex strands were not complementary at 40,874 nucleotide positions; these disagreeing positions,

representing likely sequencing or PCR errors, were removed from analysis. Again considering positions with at least 20x coverage and <5% of reads differing from reference, 3.0 million nucleotides of sequence data and 157 mutations were obtained, with an overall mutation frequency of 5.33×10^{-5}, indicative of removal of >98% of mutations seen in raw analysis and thereby demonstrating the capability of DCS to lower the error rate of DNA sequencing.

[00118] To compare this result to the method of Kinde et al. [36], reads were grouped into families by SMI tag as before but filtered for families with a minimum of 3 members. This resulted in 1.4 million nucleotides of sequence data and 61 mutations, with an overall mutation frequency of 4.25×10^{-5}. Thus, the method of Kinde et al., with a minimum family size of 3, resulted in less than half as much resultant sequence data after filtering than was obtained by DCS with a minimum family size of 1. Thus, DCS lowered the error rate of sequencing to a comparable degree to a method considered state-of-the-art, but with less loss of sequencing capacity.

Discussion

[00119] It was demonstrated that DCS analysis, using sheared DNA ends as unique molecular identifiers, results in a lowering of the apparent error rate of DNA sequencing. As this proof-of-concept experiment was performed on a library that was not optimized to maximize recovery of both strands, there were not sufficient strand-pairs recovered to perform DCS analysis with a minimum family size of >1 member. Requiring family sizes >1 is expected to further reduce sequencing errors. Moreover, this analysis was limited in that it did not include ligation of degenerate SMI tag sequences; owing to the limited number of shear points flanking any given nucleotide position, use of shear points as SMIs limits the number of unique molecules that can be sequenced in a single experiment. The use of shear points as SMIs in conjunction with an exogenously ligated SMI tag sequence would allow for increased depth of sequencing at any given nucleotide position.

Example 4: Demonstrations of duplex consensus sequencing

[00120] In addition to those described in Examples 1 and 2 above, the following materials and methods were also used.

Materials and Methods

[00121] *Construction of M13mp2 Variants*. M13mp2 gapped DNA encoding the LacZ α fragment was extended by human DNA polymerase δ [2] and the resultant products were transformed into Escherichia coli and subjected to blue-white color screening as previously described [3]. Mutant plaques were sequenced to determine the location of the mutation resulting in the color phenotype. A series of mutants, each differing from wild type by a single nucleotide change, were then mixed together with wild-type M13mp2 DNA to result in a single final mixture with distinct mutants represented at ratios of 1/10 (G6267A), 1/100 (T6299C), 1/1,000 (G6343A), and 1/10,000 (A6293T).

[00122] *Oxidative Damage of M13mp2 DNA*. Induction of DNA damage was performed by minor modifications to a published protocol [5]: 300 ng of M13mp2 double-stranded DNA was incubated in 10 mM sodium phosphate buffer, pH 7.0, in the presence of 10 μM iron sulfate and 10 μM freshly diluted hydrogen peroxide. Incubation proceeded for 30 min at 37 °C in open 1.5-mL plastic microcentrifuge tubes.

[00123] *DNA Isolation*. M13mp2 DNA was isolated from E. coli strain MC1061 by Qiagen Miniprep. To allow for greater sequencing depth at a defined region of the M13mp2 genome, an 840-bp fragment was enriched by complete digestion with the restriction enzymes Bsu36I and NaeI (New England Biolabs), followed by isolation of the fragment on an agarose gel by the Recochip system (Takara Bio).

Duplex Consensus Sequencing of M13 DNA removes artifactual sequencing errors.

[00124] The spontaneous mutation rate of M13mp2 DNA has been well established by a number of exquisitely sensitive genetic assays to be 3.0E-6 [53], that is, an average of one spontaneous base substitution error for every 330,000 nucleotides. Thus this substrate is well suited as a control for determining the background error frequency of DNA sequencing. M13mp2 DNA was sheared and ligated to adaptors containing double-stranded complementary SMI sequences by standard protocols, and was subjected to deep sequencing on an Illumina HiSeq 2000 followed by Consensus Sequencing analysis (Figure 6).

[00125] Analysis of the data by standard methods (i.e., without consideration of

the double stranded SMI sequences) resulted in an error frequency of 3.8E-03, more than one thousand fold higher than the true mutation frequency of M13mp2 DNA. This indicates that >99.9% of the apparent mutations identified by standard sequencing are in fact artifactual errors.

[00126] The data were then analyzed by Single Strand Consensus Sequencing (SSCS), using the unique SMI tag affixed to each molecule to group PCR products together in order to create a consensus of all PCR products that came from an individual molecule of single-stranded DNA. This resulted in a mutation frequency of 6.4E-OS, suggesting that -98% of sequencing errors are corrected by SSCS.

[00127] Next, the data were subjected to Duplex Consensus Sequencing (DCS), which further corrects errors by using the complementary SMI tags to compare the DNA sequence arising from both of the two strands of a single molecule of duplex DNA. This approach resulted in a mutation frequency of 2.SE-06, in nearly perfect agreement with the true mutation frequency of M13mp2 DNA of 3.0E-06. The number of nucleotides of DNA sequence obtained by a standard sequencing approach, and after SSCS and DCS analysis, may be found in Table 1 below.

[00128] **Table 1**. Data yield from Duplex Sequencing

	M13mp2 DNA	Mitochondrial DNA
Initial nucleotides	6.5×10^9	6.2×10^9
SSCS nucleotides	8.7×10^7	4.1×10^8
DCS nucleotides	$2.2E \times 10^7$	9.7×10^7
Initial reads per SSCS read	75	15
Initial reads per DCS read	295	64
SSCS reads per DCS read	4	4

Initial nucleotides represent raw reads that contain the expected fixed adapter sequence following 12 degenerate nucleotides and map to the reference genome. Apparent nucleotide loss in converting initial reads to SSCSs occurs because many of the initial reads intentionally represent identical PCR duplicates of single-stranded DNA molecules to allow for removal of sequencing and PCR errors by comparison of the sequence among the duplicates. A minimum of three initial reads are required to produce one SSCS; however, a greater average number is necessary to ensure that most DNA fragments have at least this number of duplicates. Under fully optimized conditions, each DCS read would arise from exactly two SSCS reads (one arising from each strand of the initial molecule of duplex DNA). An SSCS:DCS ratio greater than 2 indicates that the strand partner of some SSCSs was not recovered.

[00129] For an artifactual error to be scored by DCS, complementary artifactual errors must occur on both strands of a molecule of duplex DNA. Thus the background (artifactual) error frequency of DCS may be calculated as: (probability of error on one strand) * (probability of error on other strand) * (probability that both errors are complementary).

[00130] As the background error frequency of SSCS in this experiment was -6E-S, the background error frequency of DCS can be calculated as 6E-S * 6E-S * (1/3) = 1.2E-9. This represents a greater than 3 million fold improvement over the error rate of 3.SE-03 that was obtained by a standard sequencing approach.

Consensus sequencing reveals likely sites of DNA damage

[00131] M13mp2 DNA was sequenced as detailed above, with DCS adaptors containing double-stranded complementary SMIs. The spectrum of mutations obtained with SSCS was determined. Data was filtered to consist of forward-mapping reads from Read 1, i.e. sequencing of the reference strand, and reverse-mapping reads from Read 1, i.e. sequencing of the anti-reference strand. True mutations would result in an equal balance between mutations on the reference strand and their complementary mutation on the anti-reference strand.

[00132] However, SSCS analysis revealed a large number of single-stranded G→T mutations on reads mapping in the forward orientation to the reference genome, with a much smaller number of C→A mutations mapping to the reverse orientation. The spectrum of mutations identified by both SSCS and DCS analysis were examined relative to literature reference values [53] for the M13mp2 substrate (Figure 7A). SSCS analysis revealed a large excess of G→A/C→T and G→T/C→A mutations relative to reference ($P < 10^{-6}$, two-sample t test). In contrast, DCS analysis was in excellent agreement with the literature values with the exception of a decrease relative to reference of these same mutational events: G→A/C→T and G→T/C→A ($P < 0.01$). To probe the potential cause of these spectrum deviations, the SSCS data were filtered to consist of forward-mapping reads from read 1 (i.e., direct sequencing of the reference strand) and the reverse complement of reverse-mapping reads from read 1 (i.e., direct sequencing of the antireference strand.) True double-stranded mutations should result in an equal balance of complementary mutations observed on the reference and

antireference strand. However, SSCS analysis revealed a large number of single-stranded G→T mutations, with a much smaller number of C→A mutations (Figure 7B). A similar bias was seen with a large excess of C→T mutations relative to G→A mutations.

[00133] Base-specific mutagenic DNA damage is a likely explanation of these imbalances. Excess G→T mutations are consistent with the oxidative product 8-oxo-guanine (8-oxo-G) causing first round PCR errors and artifactual G→T mutations. DNA polymerases, including those commonly used in PCR, have a strong tendency to insert adenine opposite 8-oxo-G [45, 54], and misinsertion of A opposite 8-oxo-G would result in erroneous scoring of a G→T mutation. Likewise, the excess C→T mutations are consistent with spontaneous deamination of cytosine to uracil [47], a particularly common DNA damage event that results in insertion during PCR of adenine opposite uracil and erroneous scoring of a C→T mutation.

[00134] To determine whether the excess G→T mutations seen in SSCSs might reflect oxidative DNA damage at guanine nucleotides, before sequencing library preparation M13mp2 DNA was incubated with the free radical generator hydrogen peroxide in the presence of iron, a protocol that induces DNA damage [55]. This treatment resulted in a substantial further increase in G→T mutations by SSCS analysis (Figure 8A), consistent with PCR errors at sites of DNA damage as the likely mechanism of this biased mutation spectrum. In contrast, induction of oxidative damage did not alter the mutation spectrum seen with DCS analysis (Figure 8B), indicating that duplex consensus sequences are not similarly susceptible to DNA damage artifacts.

[00135] Furthermore, relative to the literature reference values, DCS analysis results in a lower frequency of G→T/C→A and C→T/G→A mutations (Figure 7A), which are the same mutations elevated in SSCS analysis as a probable result of DNA damage. Notably, the M13mp2 LacZ assay, from which reference values have been derived, is dependent upon bacterial replication of a single molecule of M13mp2 DNA. Thus, the presence of oxidative damage within this substrate could cause an analogous first-round replication error by Escherichia coli, converting a single-stranded damage event into a fixed, double-stranded mutation during replication. The slight reduction in the frequency of these two types of mutations measured by DCS analysis may,

therefore, reflect the absence of damage-induced errors that are scored by the in vivo LacZ assay.

Consensus Sequencing accurately recovers spiked-in control mutations.

[00136] A series of M13mp2 variants were constructed which contain known single base substitutions. These variants were then mixed together at known ratios, and the mixture was prepared for sequencing with DCS adaptors with double-stranded complementary SMIs and was sequenced on an Illumina HiSeq 2000. The data was then analyzed by consensus sequencing (Figure 9). With conventional analysis of the data (i.e. without consideration of the SMI tags), variants present at a level of <1/100 could not be accurately identified. This limitation occurs because at any given position, artifactual mutations are seen at a level of nearly 1/100.

[00137] In contrast, when the data is analyzed by Single Strand Consensus Sequencing (SSCS) with -20,000 fold depth, accurate recovery of mutant sequence is seen down to one mutant molecule per 10,000 wild type molecules. Duplex Consensus Sequencing (DCS), which was not performed on this sample, would allow for detection of even rarer mutations.

REFERENCES

The references, patents and published patent applications listed below, and all references cited in the specification above are hereby incorporated by reference in their entirety, as if fully set forth herein.

[1] Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31-46.

[2] Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135-45.

[3] Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, et al. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc Natl Acad Sci USA. 2011;108:13177-82.

[4] Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature. 2011;480:368-71.

[5] García-Garcerà M, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, et al. Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. PLoS ONE. 2011;6:e24161.

[6] Fordyce SL, Ávila-Arcos MC, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen FT, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. BioTechniques. 2011;51:127-33.

[7] Druley TE, Vallania FLM, Wegner DJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA. Nat Methods. 2009;6:263-5.

[8] Out AA, van Minderhout IJHM, Goeman JJ, Ariyurek Y, Ossowski S, Schneeberger K, et al. Deep sequencing to reveal new variants in pooled DNA samples. Hum Mutat. 2009;30:1703-12.

[9] Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA. 2008;105:16266-71.

[10] Chiu RWK, Akolekar R, Zheng YWL, Leung TY, Sun H, Chan KCA, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011;342:c7401.

[11] Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008;105:10513-8.

[12] Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2105;481:506-9.

[13] Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. Science Translational Medicine. 2009;1:12ra23-12ra23.

[14] Hyman RW, Herndon CN, Jiang H, Palm C, Fukushima M, Bernstein D, et al. The dynamics of the vaginal microbiome during infertility therapy with in vitro fertilization-embryo transfer. J Assist Reprod Genet. 2012;29:105-15.

[15] LaTuga MS, Ellis JC, Cotton CM, Goldberg RN, Wynn JL, Jackson RB, et al. Beyond bacteria: a study of the enteric microbial consortium in extremely low birth weight infants. PLoS ONE. 2011;6:e27858.

[16] Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: interindividual variation and dynamic response to diet. Genome Res. 2011;21:1616-25.

[17] Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011;49:3463-9.

[18] Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS ONE. 2011;6:e24907.

[19] Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci USA. 2008;105:13081-6.

[20] De Grassi A, Segala C, Iannelli F, Volorio S, Bertario L, Radice P, et al. Ultradeep Sequencing of a Human Ultraconserved Region Reveals Somatic and Constitutional Genomic Instability. PLoS Biol. 2010;8:e1000275.

[21] Zagordi O, Klein R, Däumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Research. 2010;38:7400-9.

[22] Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17:1195-201.

[23] Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, et al. Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat Methods. 2012;9:78-80.

[24] He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464:610-4.

[25] Ameur A, Stewart JB, Freyer C, Hagström E, Ingman M, Larsson N-G, et al. Ultra-Deep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. PLoS Genet. 2011;7:e1002028.

[26] Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng. 2003;96:317-23.

[27] Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. Nucleic Acids Research. 1990;18:1687-91.

[28] Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 2008;5:1005-10.

[29] Salk J, Fox E, Loeb L. Mutational heterogeneity in human cancers: origin and consequences. Annual Review of Pathology. 2009;5:51-75.

[30] Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods. 2009;6:291-5.

[31] Vandenbroucke I, Van Marck H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques. 2011;51:167-77.

[32] Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, et al. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic Acids Research. 2012;40:e2-e.

[33] Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20:273-80.

[34] Miner BE, Stöger RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Research. 2004;32:e135.

[35] McCloskey ML, Stöger R, Hansen RS, Laird CD. Encoding PCR products with batch-stamps and barcodes. Biochem Genet. 2007;45:761-7.

[36] Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci USA. 2011;108:9530-5.

[37] Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA. 2011;108:20166-71.

[38] Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011;9:72-4.

[39] Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. Nucleic Acids Research. 2011;39:e81-e.

[40] Shiroguchi K, Jia TZ, Sims PA, Xie XS. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. Proc Natl Acad Sci USA. 2012;109:1347-52.

[41] Fu GK, Hu J, Wang P-H, Fodor SPA. Counting individual DNA molecules by the stochastic attachment of diverse labels. Proc Natl Acad Sci USA. 2011;108:9026-31.

[42] Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. Embryonic stem cells and somatic cells differ in mutation frequency and type. Proc Natl Acad Sci USA. 2002;99:3586-90.

[43] Lindahl T, Wood RD. Quality control by DNA repair. Science. 1999;286:1897-1905.

[44] Kunkel, TA. Mutational specificity of depurination. Proc Natl Acad Sci USA. 1984;81:1494-98.

[45] Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature. 1991;349:431-4.

[46] Stiller M, Green RE, Ronan M, Simons JF, Du L, He W., et al. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Acad Sci USA. 2006;103:13578-84.

[47] Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obsbet Gynecol. 2011;204:205e1-11.

[48] Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch M, D'Ascenzo M, et al. Whole exome capture in solution with 3 Gbp of data. Genome Biol. 2010;11:R62:1-8.

[49] Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010; 38:159e1---8.

[50] Kaur M, Makrigiorgos GM. Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. Nucleic Acids Res. 2003; 31:26e1---7.

[51] Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature 461, 814–818.

[52] Lynch AM, Sasaki jC, Elespuru R, Jacobson-Kram D, Thybaud V, et al. New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen. 2011; 52(3):205-23.

[53] Thomas DC, Roberts JD, Sabatino RD, Myers TW, et al. Fidelity of mammalian DNA replication and replicative DNA polymerases. Biochemistry. 1991; 30:11751-9.

[54] Kasai H, et al.(1993) Formation, inhibition of formation, and repair of oxidative 8-hydroxyguanine DNA damage. Basic Life Sci 61:257–262.

[55] McBride TJ, Preston BD, Loeb LA (1991) Mutagenic spectrum resulting from DNA damage by oxygen radicals. Biochemistry 30:207–213.

[56] Greaves LC, et al. (2009) Quantification of mitochondrial DNA mutation load. Aging Cell 8:566–572.

[57] Kraytsberg Y, Nicholas A, Caro P, Khrapko K (2008) Single molecule PCR in mtDNA mutational analysis: Genuine mutations vs. damage bypass-derived artifacts. Methods 46:269–273.

[58] Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501–509.

[59] Parsons TJ, et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368.

[60] Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032.

[61] Kennedy SR, Loeb LA, Herr AJ (2011) Somatic mutations in aging, cancer and neurodegeneration. Mech Ageing Dev.

[62] Vermulst M, et al. (2007) Mitochondrial point mutations do not limit the natural lifespan of mice. Nat Genet 39:540–543.

[63] Song S, et al. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. Proc Natl Acad Sci USA 102:4990–4995.

[64] Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Hum Mol Genet 11:133–145.

[65] Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc Natl Acad Sci USA 93:12382–12387.

[66] Khaidakov M, Heflich RH, Manjanatha MG, Myers MB, Aidoo A (2003) Accumulation of point mutations in mitochondrial DNA of aging mice. Mutat Res 526:1–7.

[67] Haag-Liautard C, et al. (2008) Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Biol 6:e204.

CLAIMS

What is claimed is

1. A single molecule identifier adaptor molecule for use in sequencing a double-stranded target nucleic acid molecule comprising

a single molecule identifier (SMI) sequence, the SMI sequence comprising at least one degenerate or semi-degenerate nucleic acid sequence; and

an SMI ligation adaptor that allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence.

2. The single molecule identifier adaptor molecule of claim 1, wherein the SMI sequence is single-stranded.

3. The single molecule identifier adaptor molecule of claim 1, wherein the SMI sequence is double-stranded.

4. The single molecule identifier adaptor molecule of claim 1, wherein the double-stranded target nucleic acid molecule is a double-stranded DNA or RNA molecule.

5. The single molecule identifier adaptor molecule of claim 1, further comprising at least two PCR primer binding sites, or at least two sequencing primer binding sites, or both.

6. The single molecule identifier adaptor molecule of claim 1, further comprising a double-stranded fixed reference sequence.

7. The single molecule identifier adaptor molecule of claim 2, wherein the degenerate or semi-degenerate nucleic acid sequence comprises a first nucleotide n-mer sequence that is between approximately 3 and 20 nucleotides in length.

8. The single molecule identifier adaptor molecule of claim 7, wherein the first nucleotide n-mer sequence is a degenerate sequence.

9. The single molecule identifier adaptor molecule of claim 3, wherein the degenerate or semi-degenerate nucleic acid sequence comprises a first nucleotide n-mer sequence, a second n-mer sequence that is complementary to the first nucleotide

n-mer sequence.

10. The single molecule identifier adaptor molecule of claim 9, wherein the first n-mer sequence comprises a nucleotide sequence that is between approximately 3 and 20 nucleotides in length.

11. The single molecule identifier adaptor molecule of claim 9, wherein the first nucleotide n-mer sequence is a degenerate sequence.

12. The single molecule identifier adaptor molecule of claim 1, wherein the degenerate or semi-degenerate DNA sequence comprises a randomly fragmented double stranded nucleic acid derived from an alternative source.

13. The single molecule identifier adaptor molecule of claim 1, wherein the SMI ligation adaptor is selected from a T-overhang, an A-overhang, a CG overhang, a blunt end, or any other ligatable nucleic acid sequence.

14. The single molecule identifier adaptor molecule of claim 1, wherein the SMI adaptor molecule is Y-shaped, U-shaped, or a combination thereof.

15. A method of obtaining the sequence of a double-stranded target nucleic acid comprising

ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex;

amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products; and

sequencing the amplified SMI-target nucleic acid products.

16. The single molecule identifier adaptor molecule of claim 15, wherein the double-stranded target nucleic acid molecule is a double-stranded DNA or RNA molecule.

17. The method of claim 15, further comprising generating an error-corrected double-stranded consensus sequence by (i) grouping the sequenced SMI-target nucleic acid products into families of paired target nucleic acid strands based on a common set of SMI sequences; and (ii) removing paired target nucleic acid strands having one or

more nucleotide positions where the paired target nucleic acid strands disagree, or alternatively removing nucleotide positions from nucleic acid strands where the paired strands disagree at that specific position.

18. The method of claim 15, wherein the double-stranded target nucleic acid molecule is a sheared double-stranded DNA or RNA fragment.

19. The method of claim 18, wherein the sheared double-stranded nucleic acid fragment further comprises a double-stranded target nucleic acid sequence ligation adaptor.

20. The method of claim 19, wherein the double-stranded target nucleic acid sequence ligation adaptor is selected from a T-overhang, an A-overhang, a CG overhang, a blunt end, or any ligatable nucleic acid sequence

21. The method of claim 20, wherein each end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule.

22. The method of claim 21, wherein each SMI adaptor molecules comprises

a single molecule identifier (SMI) sequence, the SMI sequence comprising a degenerate or semi-degenerate nucleic acid sequence; and

an SMI ligation adaptor that allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence.

23. The method of claim 22, wherein the SMI sequence is single-stranded.

24. The method of claim 22, wherein the SMI sequence is double-stranded.

25. The method of claim 23, wherein the degenerate or semi-degenerate nucleic acid sequence comprises a degenerate nucleotide n-mer sequence.

26. The method of claim 24, wherein the double-stranded degenerate or semi-degenerate nucleic acid sequence comprises a first nucleotide n-mer sequence and a second n-mer sequence that is complementary to the first nucleotide n-mer sequence.

27. The method of claim 26, wherein the first nucleotide n-mer sequence is a degenerate sequence.

28. The method of claim 22, further comprising at least two PCR primer binding sites, at least two sequencing primer binding sites, a double-stranded fixed reference sequence or a combination thereof.

29. A method of generating an error corrected double-stranded consensus sequence comprising a first stage of single strand consensus sequencing (SSCS) which comprises:

tagging an individual duplex DNA molecule with an SMI adaptor molecule;

generating a set of PCR duplicates of the tagged DNA molecule by performing PCR; and

creating a single strand consensus sequence from all of the PCR duplicates which arose from an individual molecule of single-stranded DNA, resulting in two single strand consensus sequences.

30. The method of claim 29, further comprising a second stage of duplex consensus sequencing (DCS) which comprises

comparing the sequence of the two single strand consensus sequences arising from a single duplex DNA molecule; and

reducing sequencing or PCR errors by (i) grouping the sequenced SMI-target nucleic acid products into families of paired target nucleic acid strands based on a common set of SMI sequences; and (ii) removing paired target nucleic acid strands having one or more nucleotide positions where the paired target nucleic acid strands disagree, or alternatively removing nucleotide positions from nucleic acid strands where the paired strands disagree at that specific position

31. The method of claim 29, wherein the SMI adaptor molecules comprises

a single molecule identifier (SMI) sequence, the SMI sequence comprising a degenerate or semi-degenerate nucleic acid sequence; and

an SMI ligation adaptor that allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence.

32. The method of claim 31, wherein the SMI sequence is single-stranded.

33. The method of claim 31, wherein the SMI sequence is double-stranded.

34. The method of claim 32, wherein the degenerate or semi-degenerate nucleic acid sequence comprises a degenerate nucleotide n-mer sequence.

35. The method of claim 33, wherein the double-stranded degenerate or semi-degenerate nucleic acid sequence comprises a first nucleotide n-mer sequence and a second n-mer sequence that is complementary to the first nucleotide n-mer sequence.

36. The method of claim 35, wherein the first nucleotide n-mer sequence is a degenerate sequence.

Figure 1

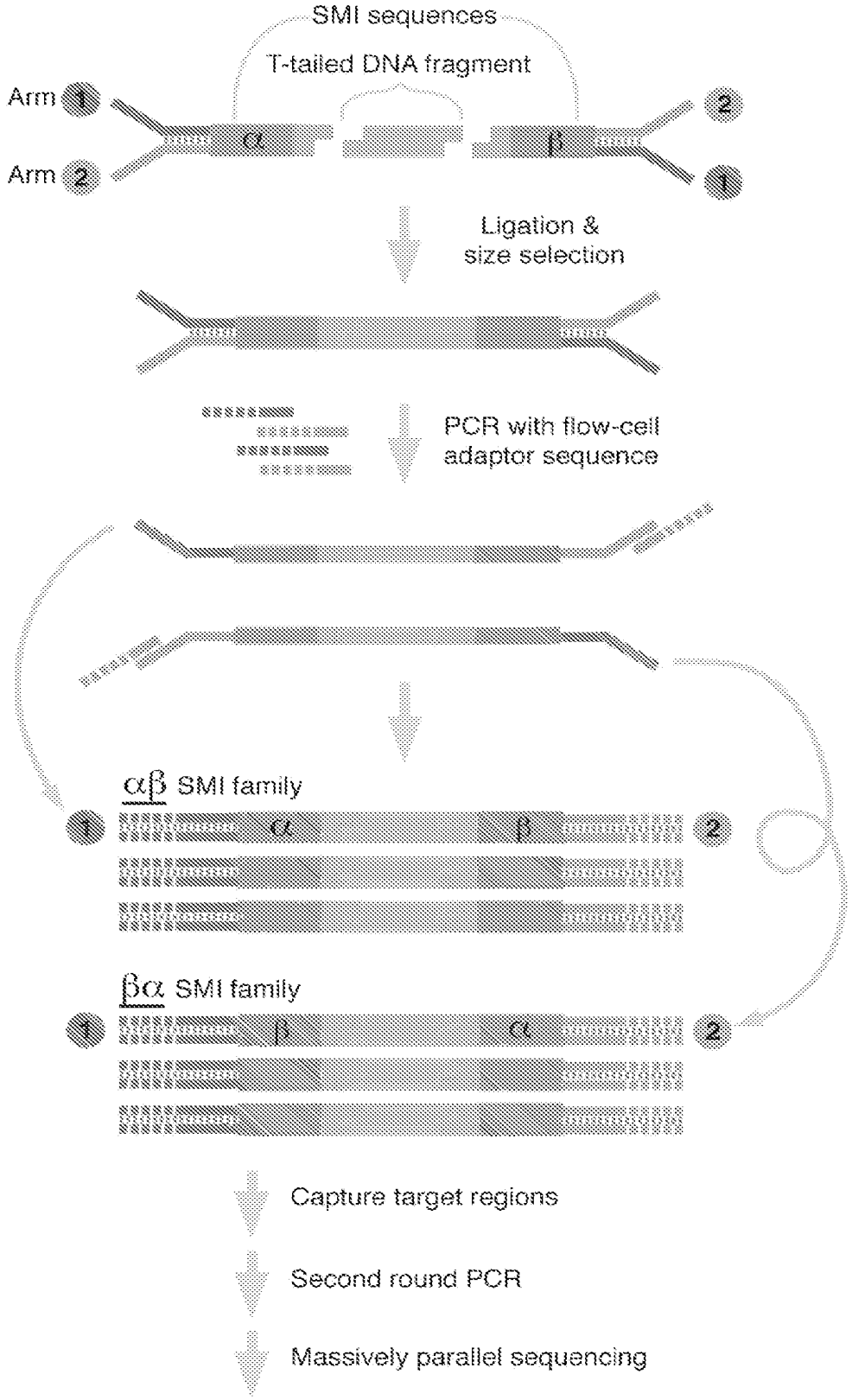

SMI sequences
T-tailed DNA fragment
Arm 1
Arm 2
α
β
2
1
Ligation &
size selection
PCR with flow-cell
adaptor sequence
αβ SMI family
βα SMI family
Capture target regions
Second round PCR
Massively parallel sequencing

Figure 2

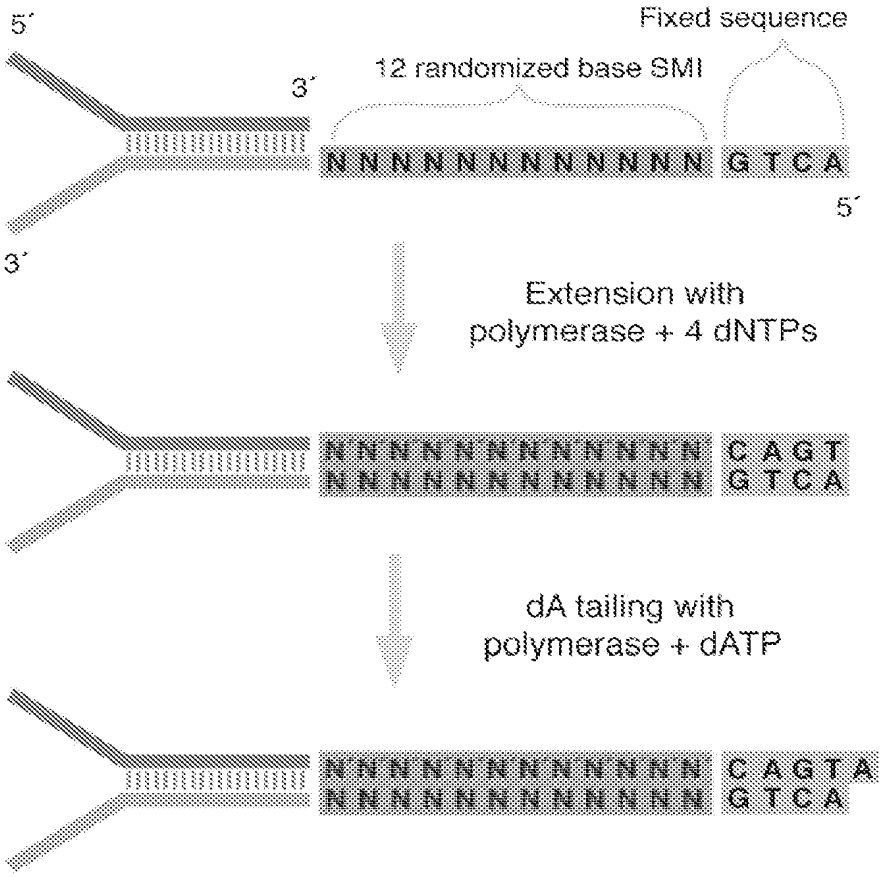

5´
3´
12 randomized base SMI
Fixed sequence
N N N N N N N N N N N N
G T C A
5´
3´
Extension with
polymerase + 4 dNTPs
N N N N N N N N N N N N
N N N N N N N N N N N N
C A G T
G T C A
dA tailing with
polymerase + dATP
N N N N N N N N N N N N
N N N N N N N N N N N N
C A G T A
G T C A

Figure 3

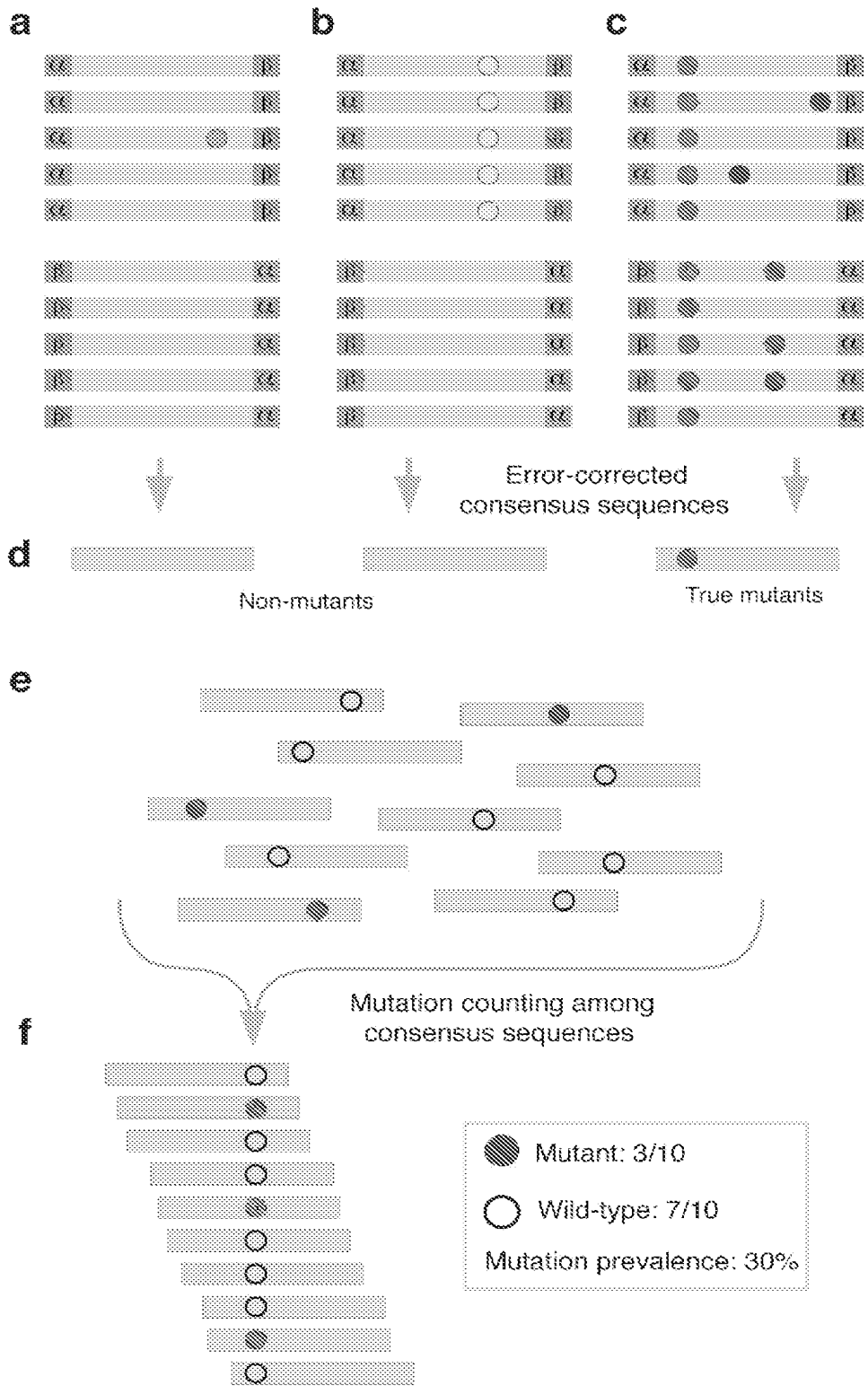

a
b
c
Error-corrected
consensus sequences
d
Non-mutants
True mutants
e
Mutation counting among
consensus sequences
f
Mutant: 3/10
Wild-type: 7/10
Mutation prevalence: 30%

Figure 4

A

5’ 1-TAAC----------TCCG-2 3’ (top strand)
3’ 2-ATTG----------AGGC-1 5’ (bottom strand)

B

5’ 1-TAAC----------TCCG-2 3’ (top strand)

5’ 1-CGGA----------GTTA-2 3’ (bottom strand)

C

1-TAAC---x----------TCCG-2 (top strand)
2-ATTG---y----------AGGC-1 (bottom strand)

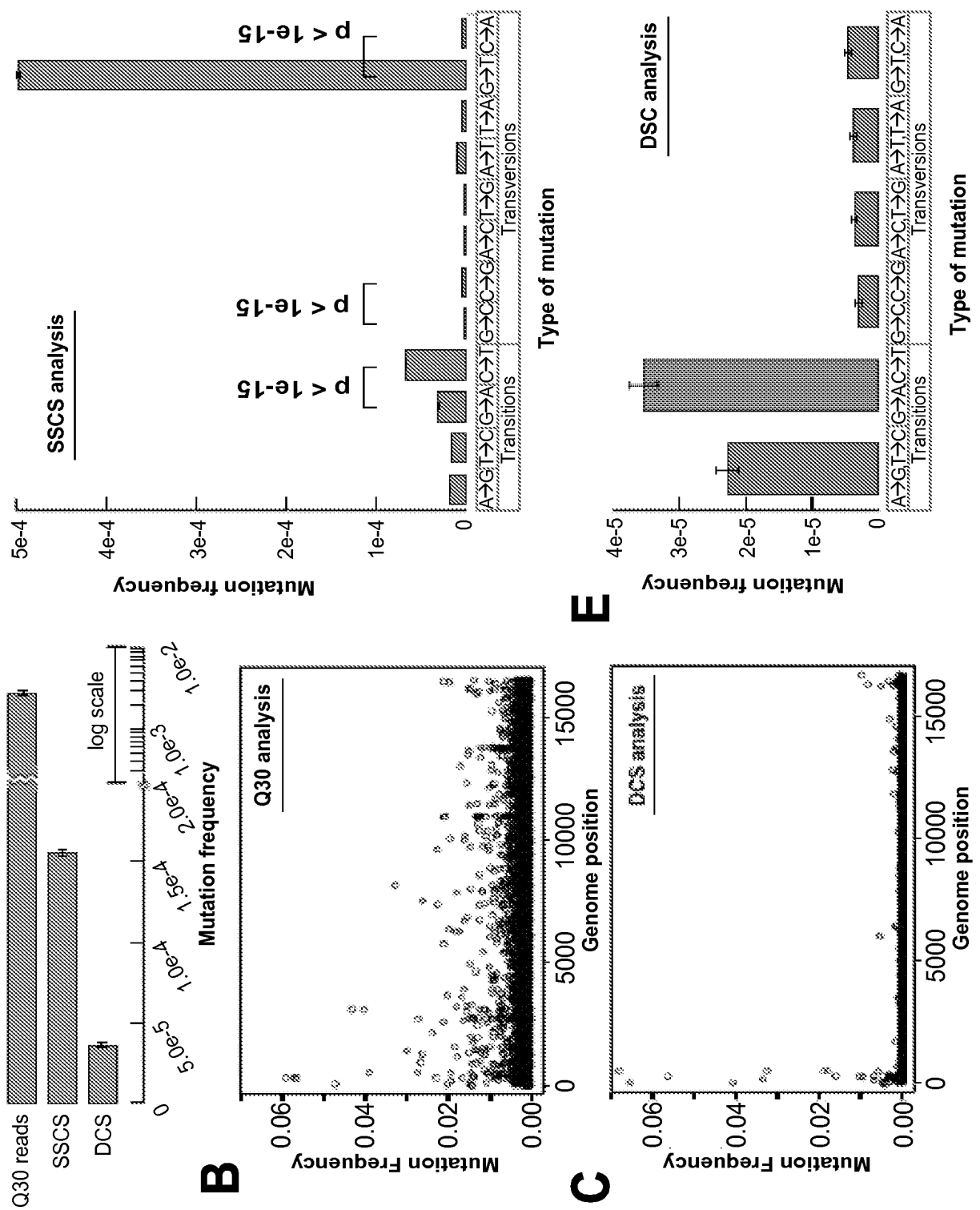

A
Q30 reads
SSCS
DCS
log scale
Mutation frequency
B
Q30 analysis
Mutation Frequency
Genome position
C
DCS analysis
Mutation Frequency
Genome position
D
SSCS analysis
p < 1e-15
p < 1e-15
p < 1e-15
Mutation frequency
Transitions
Transversions
Type of mutation
E
DSC analysis
Mutation frequency
Transitions
Transversions
Type of mutation

Figure 5

Figure 6

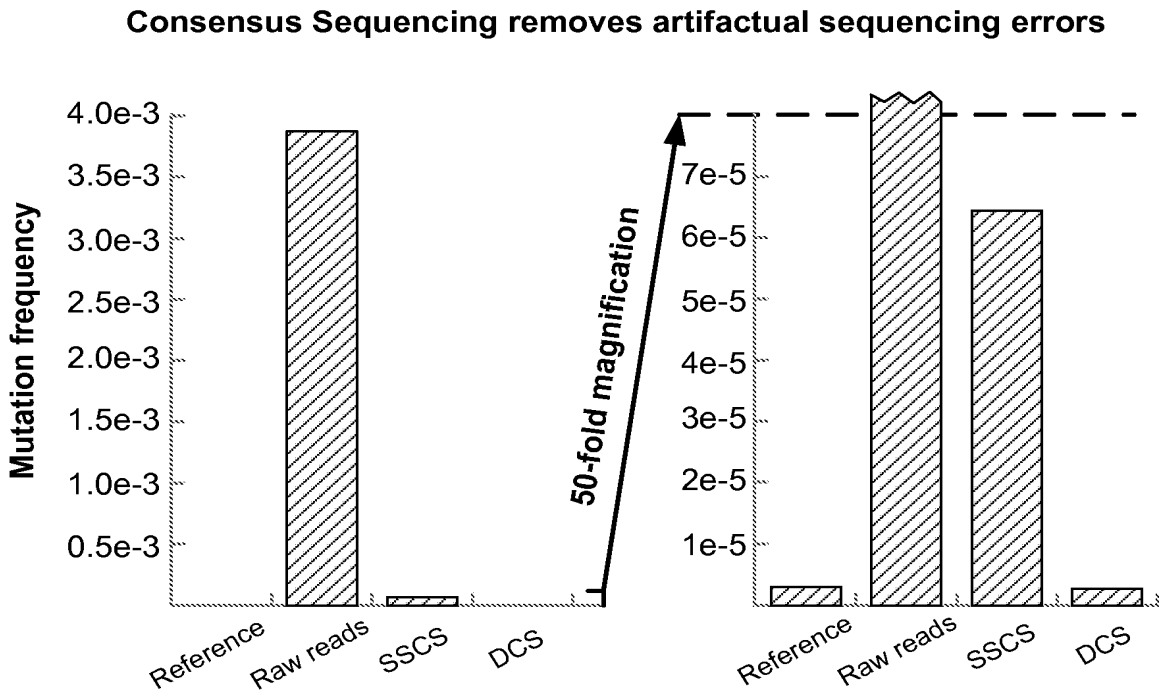

Consensus Sequencing removes artifactual sequencing errors
Mutation frequency
4.0e-3
3.5e-3
3.0e-3
2.5e-3
2.0e-3
1.5e-3
1.0e-3
0.5e-3
50-fold magnification
7e-5
6e-5
5e-5
4e-5
3e-5
2e-5
1e-5
Reference
Raw reads
SSCS
DCS
Reference
Raw reads
SSCS
DCS

Figure 7

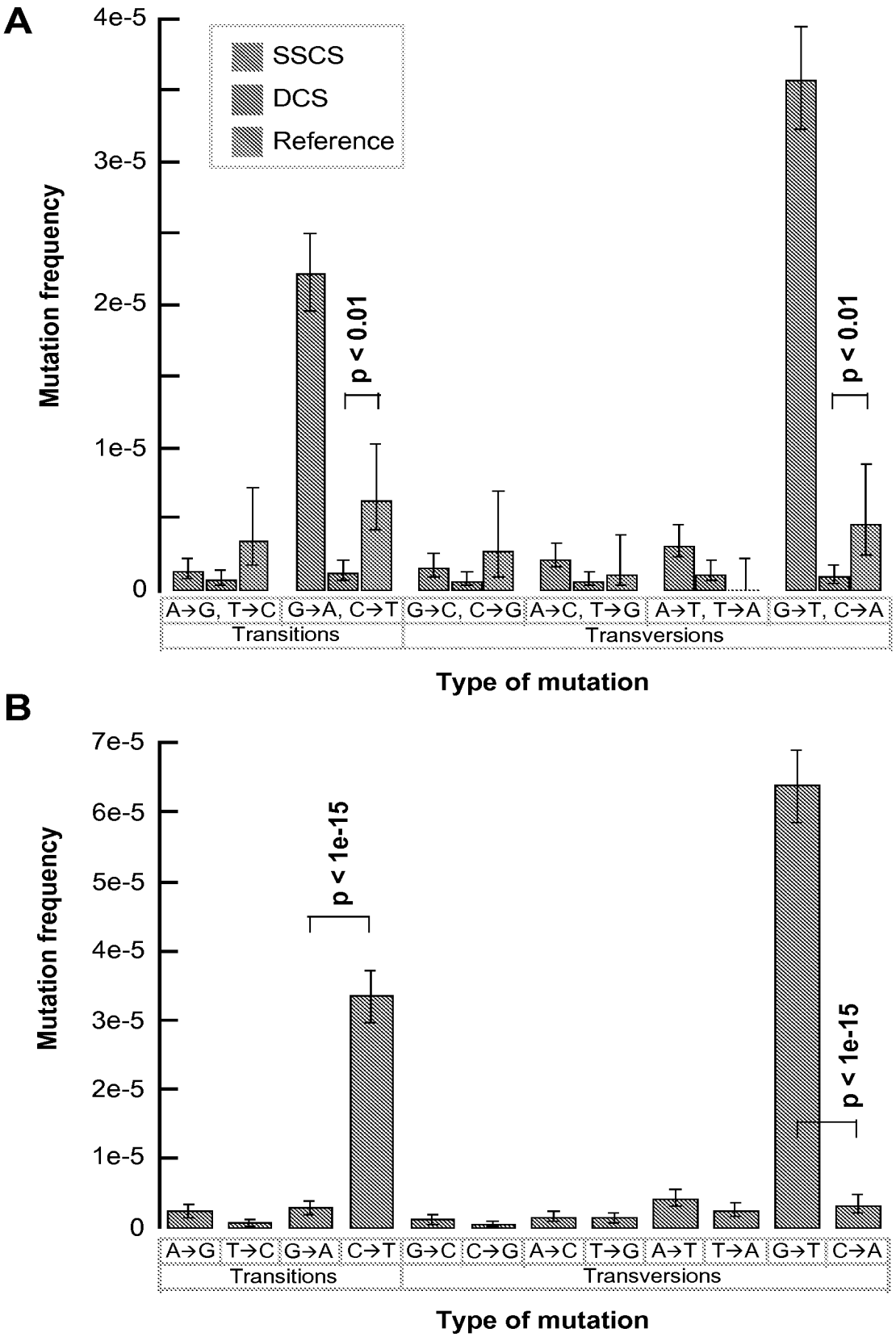

A
4e-5
3e-5
2e-5
1e-5
0
Mutation frequency
SSCS
DCS
Reference
p < 0.01
p < 0.01
A→G, T→C
G→A, C→T
G→C, C→G
A→C, T→G
A→T, T→A
G→T, C→A
Transitions
Transversions
Type of mutation
B
7e-5
6e-5
5e-5
4e-5
3e-5
2e-5
1e-5
0
Mutation frequency
p < 1e-15
p < 1e-15
A→G
T→C
G→A
C→T
G→C
C→G
A→C
T→G
A→T
T→A
G→T
C→A
Transitions
Transversions
Type of mutation

Figure 8

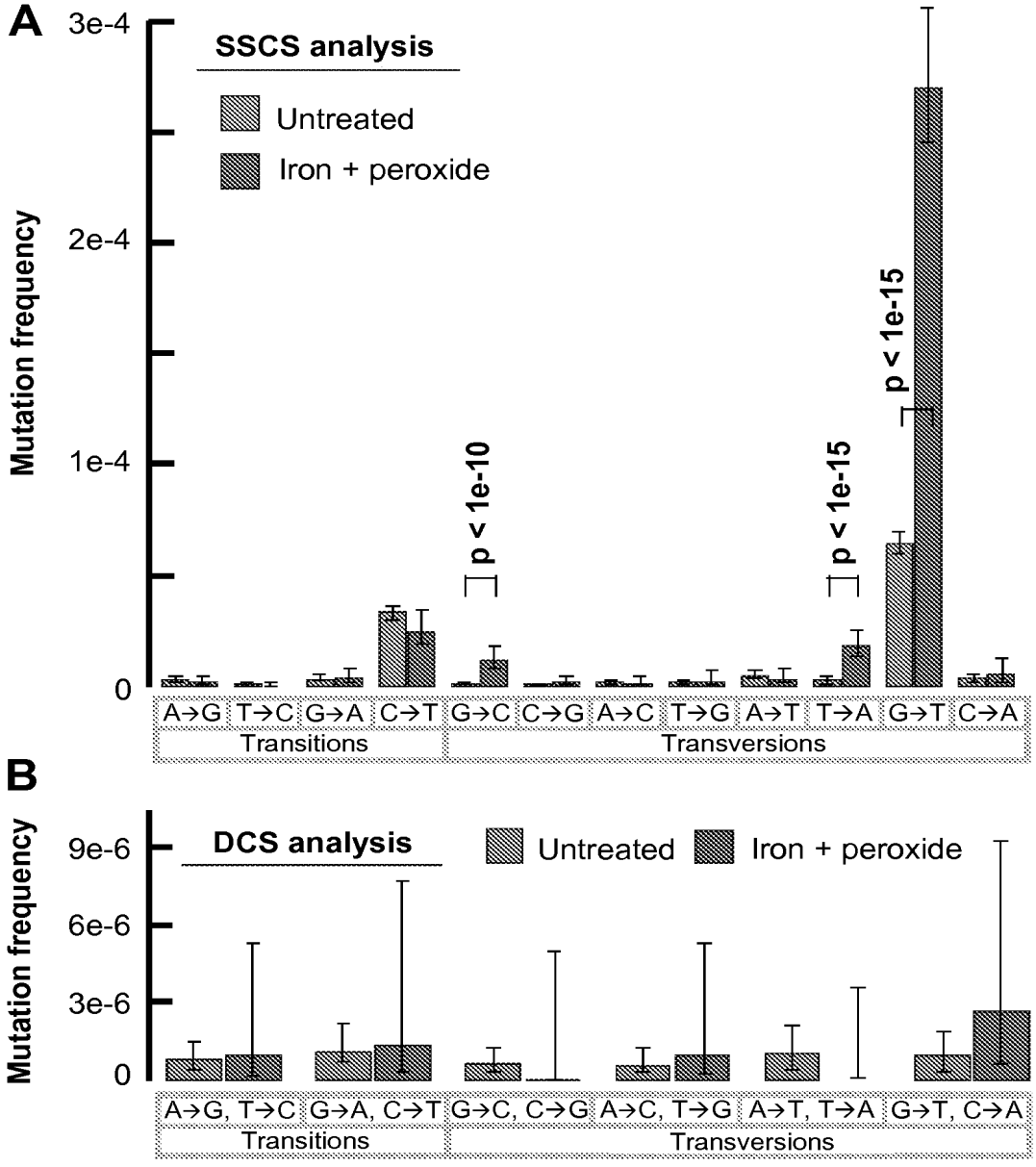

A
SSCS analysis
Untreated
Iron + peroxide
Mutation frequency
3e-4
2e-4
1e-4
0
p < 1e-10
p < 1e-15
p < 1e-15
A→G T→C G→A C→T G→C C→G A→C T→G A→T T→A G→T C→A
Transitions
Transversions
B
DCS analysis
Untreated
Iron + peroxide
Mutation frequency
9e-6
6e-6
3e-6
0
A→G, T→C G→A, C→T G→C, C→G A→C, T→G A→T, T→A G→T, C→A
Transitions
Transversions

Figure 9

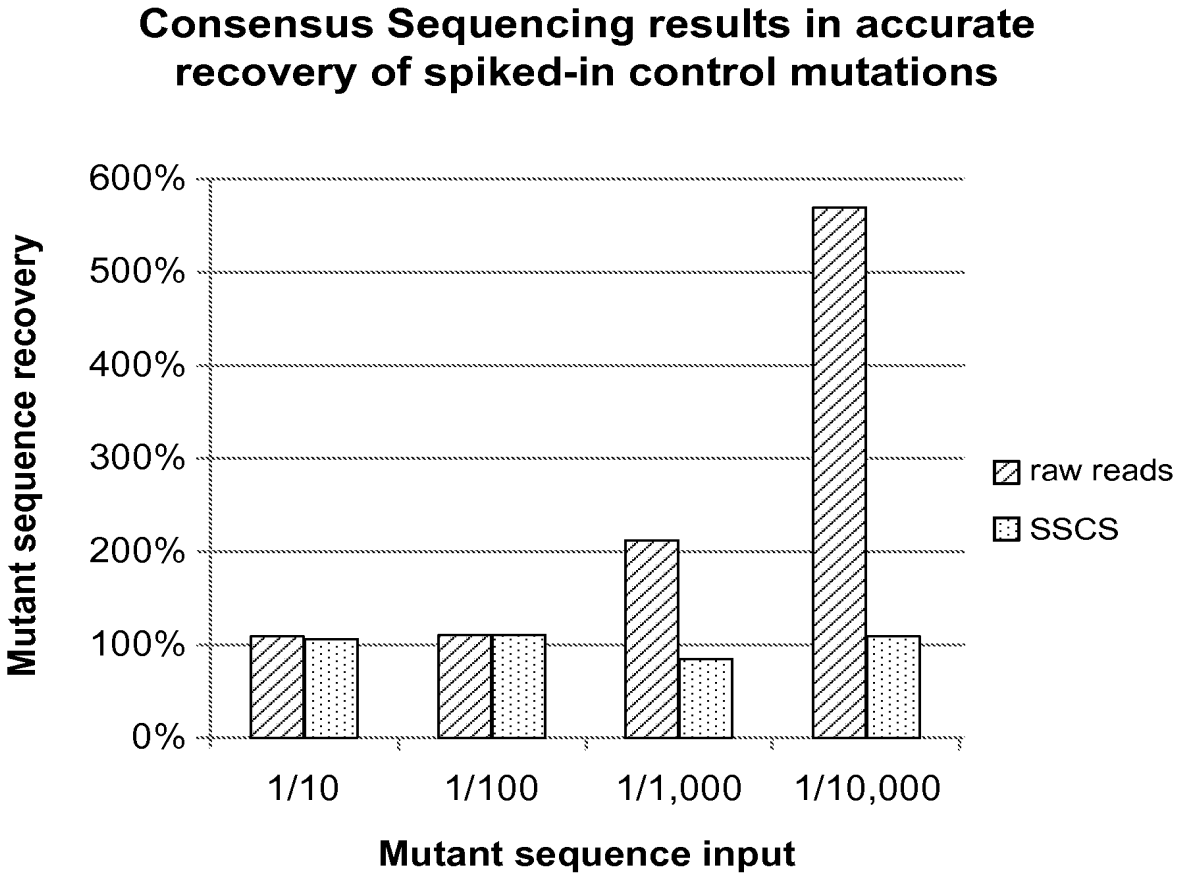

Consensus Sequencing results in accurate recovery of spiked-in control mutations
Mutant sequence recovery
600%
500%
400%
300%
200%
100%
0%
1/10
1/100
1/1,000
1/10,000
Mutant sequence input
raw reads
SSCS

Figure 10

Python code for pairing DCS reads among partner strands

```
import sys
import pysam
from optparse import OptionParser

#This program takes as input a BAM file with DCS SMI's in the header, and searches
for the partner SMI. Reads with paired SMIs are kept. Non-agreeing positions
within a read are replaced with N's.

parser=OptionParser()
parser.add_option("--infile", action="store", type='string', dest="infile",
help="input BAM file", default='sys.stdin')
parser.add_option("--outfile", action="store", type='string', dest="outfile",
help="output BAM file", default='/dev/stdout')
parser.add_option("--readnumloc", action="store", type='int', dest="readnumloc",
help="header field containing read number", default='3')
parser.add_option("--tagloc", action="store", type='int', dest="tagloc",
help="header field containing SMI", default='2')
o, args = parser.parse_args()

inBam = pysam.Samfile( o.infile, "rb" )
readDict = {}

dictctr = 0
seqctr = 0
tagmatchctr = 0
partialmatchctr = 0
seqreplacectr = 0
Nctr = 0

#first, build a dictionary with read 1 SMI's as key, and the corresponding
sequence as an entry.
for line in inBam :

     lineSplit = line.qname
     read = lineSplit.split(":")[o.readnumloc]
     tag = lineSplit.split(":")[o.tagloc]
     if read == '1' and tag not in readDict :

          readDict[tag] = [line.seq, '']
          dictctr += 1

     if dictctr % 1000000 == 0 :
          print >> sys.stderr, "sequences added to dictionary:", dictctr
          dictctr += 1

inBam.close()
inBam = pysam.Samfile( o.infile, "rb" )

#next, evaluate every read 2 SMI for a match in the dictionary

for line in inBam :

     seqctr += 1
     lineSplit = line.qname
     read = lineSplit.split(":")[o.readnumloc]
```

Figure 10 (cont.)

```
     tag = lineSplit.split(":")[o.tagloc]
     switchtag = tag[10:20] + tag[:10]

     if read == '2' and switchtag in readDict :
          tagmatchctr += 1

          if len(line.seq) == len(readDict[switchtag][0]) :
               newSeq = ''
               for i in xrange (len(line.seq) ) :

                    if line.seq[i] == readDict[switchtag][0][i] :
                         newSeq = newSeq + line.seq[i]
                    else :
                         newSeq = newSeq + 'N'


               if line.seq != readDict[switchtag][0] and newSeq.count('N') < 20 :
                    partialmatchctr += 1
                    Nctr += newSeq.count('N')

               if newSeq.count('N') < (readDict[switchtag][1]).count('N') or (
readDict[switchtag][1] == '' and newSeq.count('N') < 20 ) :
                    readDict[switchtag][1] = newSeq
                    seqreplacectr += 1

     if seqctr % 1000000 == 0 :
          print >> sys.stderr, "tags processed for matches:", seqctr
          print >> sys.stderr, "tag matches:", tagmatchctr
          print >> sys.stderr, "total sequence matches:", seqreplacectr
          print >> sys.stderr, "reads containing disagreeing bases (replaced with
N's):", partialmatchctr
          print >> sys.stderr, "number of N's added:", Nctr


inBam.close()

# Done generating tag dictionary. Reinterate over bamfile and write entries that
have a sequence match.

inBam = pysam.Samfile( o.infile, "rb" )
outBam = pysam.Samfile ( o.outfile, "wb", template=inBam)

printlinectr = 0
printlinematch = 0

for line in inBam :

     printlinectr += 1

     lineSplit = line.qname

     tag = lineSplit.split(":")[o.tagloc]
     read = lineSplit.split(":")[o.readnumloc]

     if tag in readDict and read == '1' and len (readDict[tag][1]) > 0 :

          line.seq = readDict[tag][1]
          readDict[tag][1] = ''
```

Figure 10 (cont.)

```
            printlinematch += 1
            outBam.write(line)

        if printlinectr % 1000000 == 0:
            print >> sys.stderr, "Lines evaluated for printing:", printlinectr
            print >> sys.stderr, "Matching sequences printed:", printlinematch

print >> sys.stderr, "Total tags processed for matches:", seqctr
print >> sys.stderr, "Total tag matches:", tagmatchctr
print >> sys.stderr, "Total sequence matches:", seqreplacectr
print >> sys.stderr, "Total reads containing disagreeing bases (replaced with
N's):", partialmatchctr
print >> sys.stderr, "total number of N's added:", Nctr

inBam.close()
outBam.close()
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/032665

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 20/04 (2013.01)
USPC - 506/4
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C40B 20/04; C12Q 1/68; C40B 40/06, 60/12 (2013.01)
USPC - 506/4; 435/6, 6.1; 506/16, 39

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC -B01J 2219/00596, 2219/00585, 2219/00659 (2013.01)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Orbit, Google Patent, Google Scholar

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2011/0301042 A1 (STEINMANN et al) 08 Decembr 2011 (08.12.2011) entire document	15, 16, 18-21 --- 1-14, 17, 22-36
Y	US 2008/0167195 A1 (LI et al) 10 July 2008 (10.07.2008) entire document	1-14, 22-28, 31-36
Y	US 2009/0298075 A1 (TRAVERS et al) 03 December 2009 (03.12.2009) entire document	17, 29-36

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
26 June 2013	09 JUL 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

SEQUENCE LISTING

<110> UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION

<120> METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

<130> 72227.8043.WO00

<140> PCT/US2013/032665
<141> 2013-03-15

<150> US 61/613,413
<151> 2012-03-20

<150> US 61/625,623
<151> 2012-04-17

<150> US 61/625,319
<151> 2012-04-17

<160> 9

<170> PatentIn version 3.5

<210> 1
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 1
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatct 58

<210> 2
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (5)..(16)
<223> n is a, c, g, or t

<400> 2
actgnnnnnn nnnnnnagat cggaagagca cacgtctgaa ctccagtcac 50

<210> 3
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 3
aatgatacgg cgaccaccga g 21

<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 4
gtgactggag ttcagacgtg tgc 23

<210> 5
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (25)..(30)
<223> position of fixed multiplexing barcode sequence

<220>
<221> misc_feature
<222> (25)..(30)
<223> n is a, c, g, or t

<400> 5
caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgc 53

<210> 6
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 6
cagta 5

<210> 7
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature

<222> (5)..(16)
<223> n is a, c, g, or t

<400> 7
actgnnnnnn nnnnnn 16

<210> 8
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(12)
<223> n is a, c, g, or t

<400> 8
nnnnnnnnnn nncagt 16

<210> 9
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(12)
<223> n is a, c, g, or t

<400> 9
nnnnnnnnnn nncagta 17

Electronic Acknowledgement Receipt

EFS ID:	33594731
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Michael SCHMITT
Customer Number:	94991
Filer:	Yang Tang/Amy Candeloro
Filer Authorized By:	Yang Tang
Attorney Docket Number:	72227-8043.US07
Receipt Date:	30-AUG-2018
Filing Date:	
Time Stamp:	19:34:59
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	Application Data Sheet	2018-08-30_CON4_ADS_72227 8043US7.pdf	1823552 48d9d19517e1ce20d74b8440148fe626311 a1d49	no	9

Warnings:

Information:					
2		2018-08-30_CON4_Application_722278043US7.PDF	3300924 96628e003028274c53af793c158efae77fb9158a	yes	70

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Abstract	1	2
Specification	3	52
Claims	53	57
Drawings-only black and white line drawings	58	69
Other Reference-Patent/App/Search documents	70	70

Warnings:					
Information:					
3	Sequence Listing	2018-08-30_CON4_SequenceListing_PDF_722278043US7.PDF	16160 294e54091ca2e167f0594bf7f7c35bd7238826ea	no	3
Warnings:					
Information:					
		Total Files Size (in bytes):	5140636		

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.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 September 2013 (26.09.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/142389 A1

(51) International Patent Classification:
C04B 20/04 (2006.01)

(21) International Application Number:
PCT/US2013/032665

(22) International Filing Date:
15 March 2013 (15.03.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/613,413	20 March 2012 (20.03.2012)	US
61/625,623	17 April 2012 (17.04.2012)	US
61/625,319	17 April 2012 (17.04.2012)	US

(71) Applicant: **UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION** [US/US]; 4311 11th Avenue NE, Seattle, WA 98105-4608 (US).

(72) Inventors: **SCHMITT, Michael**; Seattle, WA 98105 (US). **SALK, Jesse**; Seattle, WA 98105 (US). **LOEB, Lawrence, A.**; Bellevue, WA (US).

(74) Agent: **DUEPPEN, Lara, J.**; Perkins Coie LLP, P.O. Box 1208, Seattle, WA 98111-1208 (US).

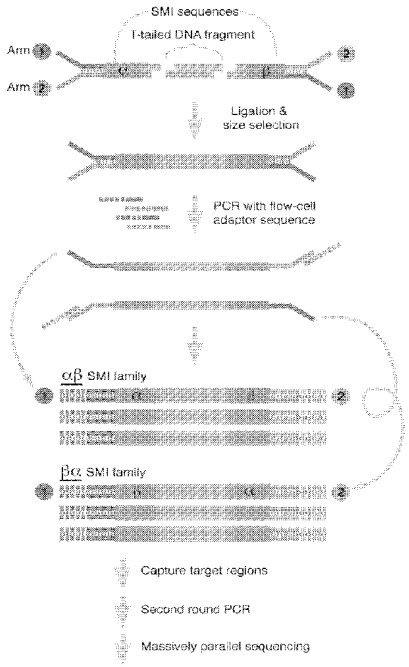
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE,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

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

[Continued on next page]

(54) Title: METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Figure 1



(57) Abstract: Next Generation DNA sequencing promises to revolutionize clinical medicine and basic research. However, while this technology has the capacity to generate hundreds of billions of nucleotides of DNA sequence in a single experiment, the error rate of approximately 1% results in hundreds of millions of sequencing mistakes. These scattered errors can be tolerated in some applications but become extremely problematic when "deep sequencing" genetically heterogeneous mixtures, such as tumors or mixed microbial populations. To overcome limitations in sequencing accuracy, a method Duplex Consensus Sequencing (DCS) is provided. This approach greatly reduces errors by independently tagging and sequencing each of the two strands of a DNA duplex. As the two strands are complementary, true mutations are found at the same position in both strands. In contrast, PCR or sequencing errors will result in errors in only one strand.

WO 2013/142389 A1

Published:

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

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

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/118,306	08/30/2018	1636	0.00	072227-8043.US07	36	3

CONFIRMATION NO. 1114

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

FILING RECEIPT

OC000000102897885

Date Mailed: 10/10/2018

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Michael Schmitt, Seattle, WA;
Jesse Salk, Seattle, WA;
Lawrence A. Loeb, Bellevue, WA;

Applicant(s)

UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION, Seattle, WA

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 15/660,785 07/26/2017
which is a CON of 14/386,800 09/20/2014 PAT 9752188
which is a 371 of PCT/US2013/032665 03/15/2013
which claims benefit of 61/613,413 03/20/2012
and claims benefit of 61/625,623 04/17/2012
and claims benefit of 61/625,319 04/17/2012

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 10/09/2018
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/118,306**
Projected Publication Date: To Be Determined - pending completion of Missing Parts
Non-Publication Request: No
Early Publication Request: No
**** SMALL ENTITY ****
Title

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 16/118,306

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	75		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	330		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	380		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	36 minus 20 =	* 16	x 50 =	800	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	3 minus 3 =	*	x 230 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	1585		TOTAL	

APPLICATION AS AMENDED - PART II

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

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

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/118,306	08/30/2018	Michael Schmitt	072227-8043.US07

CONFIRMATION NO. 1114

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

FORMALITIES LETTER

OC000000102897866

Date Mailed: 10/10/2018

NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

FILED UNDER 37 CFR 1.53(b)

Filing Date Granted

Items Required To Avoid Abandonment:

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The statutory basic filing fee is missing.
- The application search fee must be submitted.
- The application examination fee must be submitted.
- A copy of the "Sequence Listing" in computer readable form has not been submitted as required by 37 CFR 1.821(e). If the effective filing date is on or after September 8, 2000, see the final rulemaking notice published in the Federal Register at 65 FR 54604 (September 8, 2000) and 1238 OG 145 (September 19, 2000). Applicant must provide an initial computer readable form (CRF) copy of the "Sequence Listing" and a statement that the content of the sequence listing information recorded in computer readable form is identical to the written (on paper or compact disc) sequence listing and, where applicable, includes no new matter, as required by 37 CFR 1.821(e), 1.821(f), 1.821(g), 1.825(b), or 1.825(d). If applicant desires the sequence listing in the instant application to be identical with that of another application on file in the U.S. Patent and Trademark Office, such request in accordance with 37 CFR 1.821(e) may be submitted in lieu of a new CRF.

Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

For questions regarding compliance to these requirements, please contact:

- **For Rules interpretation, call (571) 272-2510**
- **For CRF submission help, call (571) 272-2510**
- **For PatentIn software program support technical assistance, call (866)217-9197.**
- **Send e-mail correspondence for PatentIn software program help to ebc@uspto.gov**
- **Patent software is available at http://www.uspto.gov/patents/resources/tools/checker/patentinrel.jsp .**

- Additional claim fees of $ **800** as a small entity, including any required multiple dependent claim fee, are required. Applicant must submit the additional claim fees or cancel the additional claims for which fees are due.
- Surcharge as set forth in 37 CFR 1.16(f) must be submitted.
  The surcharge is due for any one of:
  - late submission of the basic filing fee, search fee, or examination fee,
  - late submission of inventor's oath or declaration,
  - filing an application that does not contain at least one claim on filing, or
  - submission of an application filed by reference to a previously filed application.

SUMMARY OF FEES DUE:

The fee(s) required within **TWO MONTHS** from the date of this Notice to avoid abandonment is/are itemized below. Small entity discount is in effect. If applicant is qualified for micro entity status, an acceptable Certification of Micro Entity Status must be submitted to establish micro entity status. (See 37 CFR 1.29 and forms PTO/SB/15A and 15B.)

- $ **75** basic filing fee.
- $ **80** surcharge.
- $ **330** search fee.
- $ **380** examination fee.
- $ **800** for **16** total claims over 20.
- $(**0**) previous unapplied payment amount.
- $ **1665** TOTAL FEE BALANCE DUE.

Items Required To Avoid Processing Delays:

Applicant is notified that the above-identified application contains the deficiencies noted below. No period for reply is set forth in this notice for correction of these deficiencies. However, if a deficiency relates to the inventor's oath or declaration, the applicant must file an oath or declaration in compliance with 37 CFR 1.63, or a substitute statement in compliance with 37 CFR 1.64, executed by or with respect to each actual inventor no later than the expiration of the time period set in the "Notice of Allowability" to avoid abandonment. See 37 CFR 1.53(f).

- A properly executed inventor's oath or declaration has not been received for the following inventor(s):
  Michael Schmitt
  Jesse Salk
  Lawrence A. Loeb

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be mailed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Registered users of EFS-Web may alternatively submit their reply to this notice via EFS-Web, including a copy of this Notice and selecting the document description "Applicant response to Pre-Exam Formalities Notice". https://sportal.uspto.gov/authenticate/AuthenticateUserLocalEPF.html

For more information about EFS-Web please call the USPTO Electronic Business Center at 1-866-217-9197 or visit our website at http://www.uspto.gov/ebc.

If you are not using EFS-Web to submit your reply, you must include a copy of this notice.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/ygizaw/

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	MICHAEL SCHMITT, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1636
FILED:	AUGUST 30, 2018	EXAMINER:	
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

PRELIMINARY AMENDMENT

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

Dear Commissioner:

Prior to examination on the merits, please amend the application as follows:

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

Amendments to the claims begin on page 3 of this paper.

Remarks begin on page 10 of this paper.

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

PRIORITY CLAIM

[0001] This application is a continuation of U.S. Patent Application No. 15/660,785, filed July 26, 2017 and now pending, which is a continuation of U.S. Patent Application No. 14/386,800, filed September 20, 2014 and now U.S. Patent No. 9,752,188, which is a U.S. national stage application of International Application No. PCT/US2013/032665, filed March 15, 2013, which claims priority to U.S. Provisional Patent Application No. 61/613,413, filed March 20, 2012; U.S. Provisional Patent Application No. 61/625,623, filed April 17, 2012; and U.S. Provisional Patent Application No. 61/625,319, filed April 17, 2012; the subject matter of all of which are hereby incorporated by reference as if fully set forth herein.

STATEMENT OF GOVERNMENT INTEREST

[0002] ~~The present~~This invention was made with government support under Grant Nos. ~~RO1 CA115802 and RO1 F30AG033485,~~ R01CA102029 and R01CA115802 awarded by the National Institutes of Health. The ~~Government~~government has certain rights in the invention.

BACKGROUND

[0003] The advent of massively parallel DNA sequencing has ushered in a new era of genomic exploration by making simultaneous genotyping of hundreds of billions of base-pairs possible at small fraction of the time and cost of traditional Sanger methods [1]. Because these technologies digitally tabulate the sequence of many individual DNA fragments, unlike conventional techniques which simply report the average genotype of an aggregate collection of molecules, they offer the unique ability to detect minor variants within heterogeneous mixtures [2].

[0004] This concept of “deep sequencing” has been implemented in a variety fields including metagenomics [3, 4], paleogenomics [5], forensics [6], and human genetics [7, 8] to disentangle subpopulations in complex biological samples. Clinical

applications, such prenatal screening for fetal aneuploidy [9, 10], early detection of cancer [11] and monitoring its response to therapy [12, 13] with nucleic acid-based serum biomarkers, are rapidly being developed. Exceptional diversity within microbial [14, 15] viral [16-18] and tumor cell populations [19, 20] has been characterized through next-generation sequencing, and many low-frequency, drug-resistant variants of therapeutic importance have been so identified [12, 21, 22]. Previously unappreciated intra-organismal mosasism in both the nuclear [23] and mitochondrial [24, 25] genome has been revealed by these technologies, and such somatic heterogeneity, along with that arising within the adaptive immune system [13], may be an important factor in phenotypic variability of disease.

[0005] Deep sequencing, however, has limitations. Although, in theory, DNA subpopulations of any size should be detectable when deep sequencing a sufficient number of molecules, a practical limit of detection is imposed by errors introduced during sample preparation and sequencing. PCR amplification of heterogeneous mixtures can result in population skewing due to stoichastic and non-stoichastic amplification biases and lead to over- or under-representation of particular variants [26]. Polymerase mistakes during pre-amplification generate point mutations resulting from base mis-incorporations and rearrangements due to template switching [26, 27]. Combined with the additional errors that arise during cluster amplification, cycle sequencing and image analysis, approximately 1% of bases are incorrectly identified, depending on the specific platform and sequence context [2, 28]. This background level of artifactual heterogeneity establishes a limit below which the presence of true rare variants is obscured [29].

[0006] A variety of improvements at the level of biochemistry [30-32] and data processing [19, 21, 28, 32, 33] have been developed to improve sequencing accuracy. The ability to resolve subpopulations below 0.1%, however, has remained elusive. Although several groups have attempted to increase sensitivity of sequencing, several limitations remain. For example techniques whereby DNA fragments to be sequenced are each uniquely tagged [34, 35] prior to amplification [36-41] have been reported. Because all amplicons derived from a particular starting molecule will bear its specific tag, any variation in the sequence or copy number of identically tagged sequencing

reads can be discounted as technical error. This approach has been used to improve counting accuracy of DNA [38, 39, 41] and RNA templates [37, 38, 40] and to correct base errors arising during PCR or sequencing [36, 37, 39]. Kinde et. al. reported a reduction in error frequency of approximately 20-fold with a tagging method that is based on labeling single-stranded DNA fragments with a primer containing a 14 bp degenerate sequence. This allowed for an observed mutation frequency of ~0.001% mutations/bp in normal human genomic DNA [36]. Nevertheless, a number of highly sensitive genetic assays have indicated that the true mutation frequency in normal cells is likely to be far lower, with estimates of per-nucleotide mutation frequencies generally ranging from 10^{-9} to 10^{-11} [42]. Thus, the mutations seen in normal human genomic DNA by Kinde et al. are likely the result of significant technical artifacts.

[0007] Traditionally, next-generation sequencing platforms rely upon generation of sequence data from a single strand of DNA. As a consequence, artifactual mutations introduced during the initial rounds of PCR amplification are undetectable as errors - even with tagging techniques - if the base change is propagated to all subsequent PCR duplicates. Several types of DNA damage are highly mutagenic and may lead to this scenario. Spontaneous DNA damage arising from normal metabolic processes results in thousands of damaging events per cell per day [43]. In addition to damage from oxidative cellular processes, further DNA damage is generated *ex vivo* during tissue processing and DNA extraction [44]. These damage events can result in frequent copying errors by DNA polymerases: for example a common DNA lesion arising from oxidative damage, 8-oxo-guanine, has the propensity to incorrectly pair with adenine during complementary strand extension with an overall efficiency greater than that of correct pairing with cytosine, and thus can contribute a large frequency of artifactual G→T mutations [45]. Likewise, deamination of cytosine to form uracil is a particularly common event which leads to the inappropriate insertion of adenine during PCR, thus producing artifactual C→T mutations with a frequency approaching 100% [46].

[0008] It would be desirable to develop an approach for tag-based error correction, which reduces or eliminates artifactual mutations arising from DNA damage, PCR errors, and sequencing errors; allows rare variants in heterogeneous populations

to be detected with unprecedented sensitivity; and which capitalizes on the redundant information stored in complexed double-stranded DNA.

SUMMARY

[0009] In one embodiment, a single molecule identifier (SMI) adaptor molecule for use in sequencing a double-stranded target nucleic acid molecule is provided. Said SMI adaptor molecule includes a single molecule identifier (SMI) sequence which comprises a degenerate or semi-degenerate DNA sequence; and an SMI ligation adaptor that allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence. The SMI sequence may be single-stranded or double-stranded. In some embodiments, the double-stranded target nucleic acid molecule is a double-stranded DNA or RNA molecule.

[0010] In another embodiment, a method of obtaining the sequence of a double-stranded target nucleic acid is provided (also known as Duplex Consensus Sequencing or DCS) is provided. Such a method may include steps of ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex; amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products; and sequencing the amplified SMI-target nucleic acid products.

[0011] In some embodiments, the method may additionally include generating an error-corrected double-stranded consensus sequence by (i) grouping the sequenced SMI-target nucleic acid products into families of paired target nucleic acid strands based on a common set of SMI sequences; and (ii) removing paired target nucleic acid strands having one or more nucleotide positions where the paired target nucleic acid strands are non-complementary (or alternatively removing individual nucleotide positions in cases where the sequence at the nucleotide position under consideration disagrees among the two strands). In further embodiments, the method confirms the presence of a true mutation by (i) identifying a mutation present in the paired target nucleic acid strands having one or more nucleotide positions that disagree; (ii) comparing the mutation present in the paired target nucleic acid strands to the error corrected

double-stranded consensus sequence; and (iii) confirming the presence of a true mutation when the mutation is present on both of the target nucleic acid strands and appears in all members of a paired target nucleic acid family.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 illustrates an overview of Duplex Consensus Sequencing. Sheared double-stranded DNA that has been end-repaired and T-tailed is combined with A-tailed SMI adaptors and ligated according to one embodiment. Because every adaptor contains a unique, double-stranded, complementary n-mer random tag on each end (n-mer = 12 bp according to one embodiment), every DNA fragment becomes labeled with two distinct SMI sequences (arbitrarily designated α and β in the single capture event shown). After size-selecting for appropriate length fragments, PCR amplification with primers containing Illumina flow-cell-compatible tails is carried out to generate families of PCR duplicates. By virtue of the asymmetric nature of adapted fragments, two types of PCR products are produced from each capture event. Those derived from one strand will have the α SMI sequence adjacent to flow-cell sequence 1 and the β SMI sequence adjacent to flow cell sequence 2. PCR products originating from the complementary strand are labeled reciprocally.

[0013] Figure 2 illustrates Single Molecule Identifier (SMI) adaptor synthesis according to one embodiment. Oligonucleotides are annealed and the complement of the degenerate lower arm sequence (N's) plus adjacent fixed bases is produced by polymerase extension of the upper strand in the presence of all four dNTPs. After reaction cleanup, complete adaptor A-tailing is ensured by extended incubation with polymerase and dATP.

[0014] Figure 3 illustrates error correction through Duplex Consensus Sequencing (DCS) analysis according to one embodiment. (a-c) shows sequence reads (brown) sharing a unique set of SMI tags are grouped into paired families with members having strand identifiers in either the αβ or βα orientation. Each family pair reflects one double-stranded DNA fragment. (a) shows mutations (spots) present in only one or a few family members representing sequencing mistakes or PCR-introduced errors occurring late in amplification. (b) shows mutations occurring in many or all

members of one family in a pair representing mutations scored on only one of the two strands, which can be due to PCR errors arising during the first round of amplification such as might occur when copying across sites of mutagenic DNA damage. (c) shows true mutations (* arrow) present on both strands of a captured fragment appear in all members of a family pair. While artifactual mutations may co-occur in a family pair with a true mutation, these can be independently identified and discounted when producing (d) an error-corrected consensus sequence (i.e., single stranded consensus sequence) (+ arrow) for each duplex. (e) shows consensus sequences from all independently captured, randomly sheared fragments containing a particular genomic site are identified and (f) compared to determine the frequency of genetic variants at this locus within the sampled population.

[0015] Figure 4 illustrates an example of how a SMI sequence with n-mers of 4 nucleotides in length (4-mers) are read by Duplex Consensus Sequencing (DCS) according to some embodiments. (A) shows the 4-mers with the PCR primer binding sites (or flow cell sequences) 1 and 2 indicated at each end. (B) shows the same molecules as in (A) but with the strands separated and the lower strand now written in the 5'-3' direction. When these molecules are amplified with PCR and sequenced, they will yield the following sequence reads: The top strand will give a read 1 file of TAAC--- and a read 2 file of GCCA---. Combining the read 1 and read 2 tags will give TAACCGGA as the SMI for the top strand. The bottom strand will give a read 1 file of CGGA---- and a read 2 file of TAAC---. Combining the read 1 and read 2 tags will give CGGATAAC as the SMI for the bottom strand. (C) illustrates the orientation of paired strand mutations in DCS. In the initial DNA duplex shown in Figures 4A and 4B, a mutation "x" (which is paired to a complementary nucleotide "y") is shown on the left side of the DNA duplex. The "x" will appear in read 1, and the complementary mutation on the opposite strand, "y," will appear in read 2. Specifically, this would appear as "x" in both read 1 and read 2 data, because "y" in read 2 is read out as "x" by the sequencer owing to the nature of the sequencing primers, which generate the complementary sequence during read 2.

[0016] Figure 5 illustrates duplex sequencing of human mitochondrial DNA. (A) Overall mutation frequency as measured by a standard sequencing approach, SSCS,

and DCS. (B) Pattern of mutation in human mitochondrial DNA by a standard sequencing approach. The mutation frequency (vertical axis) is plotted for every position in the □16-kb mitochondrial genome. Due to the substantial background of technical error, no obvious mutational pattern is discernible by this method. (C) DCS analysis eliminates sequencing artifacts and reveals the true distribution of mitochondrial mutations to include a striking excess adjacent to the mtDNA origin of replication. (D) SSCS analysis yields a large excess of G→T mutations relative to complementary C→A mutations, consistent with artifacts from damaged-induced 8-oxo-G lesions during PCR. All significant ($P < 0.05$) differences between paired reciprocal mutation frequencies are noted. (E) DCS analysis removes the SSCS strand bias and reveals the true mtDNA mutational spectrum to be characterized by an excess of transitions.

[0017] Figure 6 shows that consensus sequencing removes artifactual sequencing errors as compared to Raw Reads. Duplex Consensus Sequencing (DCS) results in an approximately equal number of mutations as the reference and single strand consensus sequencing (SSCS) .

[0018] Figure 7 illustrates duplex sequencing of M13mp2 DNA. (A) Single-strand consensus sequences (SSCSs) reveal a large excess of G→A/C→T and G→T/C→A mutations, whereas duplex consensus sequences (DCSs) yield a balanced spectrum. Mutation frequencies are grouped into reciprocal mispairs, as DCS analysis only scores mutations present in both strands of duplex DNA. All significant ($P < 0.05$) differences between DCS analysis and the literature reference values are noted. (B) Complementary types of mutations should occur at approximately equal frequencies within a DNA fragment population derived from duplex molecules. However, SSCS analysis yields a 15-fold excess of G→T mutations relative to C→A mutations and an 11-fold excess of C→T mutations relative to G→A mutations. All significant ($P < 0.05$) differences between paired reciprocal mutation frequencies are noted.

[0019] Figure 8 shows the effect of DNA damage on the mutation spectrum. DNA damage was induced by incubating purified M13mp2 DNA with hydrogen peroxide and FeSO4. (A) SSCS analysis reveals a further elevation from baseline of G→T mutations, indicating these events to be the artifactual consequence of nucleotide oxidation. All

significant ($P < 0.05$) changes from baseline mutation frequencies are noted. (B) Induced DNA damage had no effect on the overall frequency or spectrum of DCS mutations.

[0020] Figure 9 shows duplex sequencing results in accurate recovery of spiked-control mutations. A series of variants of M13mp2 DNA, each harboring a known single-nucleotide substitution, were mixed in together at known ratios and the mixture was sequenced to ~20,000-fold final depth. Standard sequencing analysis cannot accurately distinguish mutants present at a ratio of less than 1/100, because artifactural mutations occurring at every position obscure the presence of less abundant true mutations, rendering apparent recovery greater than 100%. Duplex consensus sequences, in contrast, accurately identify spiked-in mutations down to the lowest tested ratio of 1/10,000.

[0021] Figure 10 is a Python Code that may used to carry out methods described herein according to one embodiment.

DETAILED DESCRIPTION

[0022] Single molecule identifier adaptors and methods for their use are provided herein. According to the embodiments described herein, a single molecule identifier (SMI) adaptor molecule is provided. Said SMI adaptor molecule is double stranded, and may include a single molecule identifier (SMI) sequence, and an SMI ligation adaptor (Figure 2). Optionally, the SMI adaptor molecule further includes at least two PCR primer binding sites, at least two sequencing primer binding sites, or both.

[0023] The SMI adaptor molecule may form a "Y-shape" or a "hairpin shape." In some embodiments, the SMI adaptor molecule is a "Y-shaped" adaptor, which allows both strands to be independently amplified by a PCR method prior to sequencing because both the top and bottom strands have binding sites for PCR primers FC1 and FC2 as shown in the examples below. A schematic of a Y-shaped SMI adaptor molecule is also shown in Figure 2. A Y-shaped SMI adaptor requires successful amplification and recovery of both strands of the SMI adaptor molecule. In one embodiment, a modification that would simplify consistent recovery of both strands entails ligation of a Y-shaped SMI adaptor molecule to one end of a DNA duplex

molecule, and ligation of a "U-shaped" linker to the other end of the molecule. PCR amplification of the hairpin-shaped product will then yield a linear fragment with flow cell sequences on either end. Distinct PCR primer binding sites (or flow cell sequences FC1 and FC2) will flank the DNA sequence corresponding to each of the two SMI adaptor molecule strands, and a given sequence seen in Read 1 will then have the sequence corresponding to the complementary DNA duplex strand seen in Read 2. Mutations are scored only if they are seen on both ends of the molecule (corresponding to each strand of the original double-stranded fragment), i.e. at the same position in both Read 1 and Read 2. This design may be accomplished as described in the examples relating to double stranded SMI sequence tags.

[0024] In other embodiments, the SMI adaptor molecule is a "hairpin" shaped (or "U-shaped") adaptor. A hairpin DNA product can be used for error correction, as this product contains both of the two DNA strands. Such an approach allows for reduction of a given sequencing error rate N to a lower rate of N*N*(1/3), as independent sequencing errors would need to occur on both strands, and the same error among all three possible base substitutions would need to occur on both strands. For example, the error rate of 1/100 in the case of Illumina sequencing [32] would be reduced to (1/100)*(1/100)*(1/3) = 1/30,000.

[0025] An additional, more remarkable reduction in errors can be obtained by inclusion of a single-stranded SMI in either the hairpin adaptor or the "Y-shaped" adaptor will also function to label both of the two DNA strands. Amplification of hairpin-shaped DNA may be difficult as the polymerase must synthesize through a product containing significant regions of self-complementarity, however, amplification of hairpin-shaped structures has already been established in the technique of hairpin PCR, as described below. Amplification using hairpin PCR is further described in detail in U.S. Patent No. 7,452,699, the subject matter of which is hereby incorporated by reference as if fully set forth herein.

[0026] According to the embodiments described herein, the SMI sequence (or "tag") may be a double-stranded, complementary SMI sequence or a single-stranded SMI sequence. In some embodiments, the SMI adaptor molecule includes an SMI

sequence (or "tag") of nucleotides that is degenerate or semi-degenerate. In some embodiments, the degenerate or semi-degenerate SMI sequence may be a random degenerate sequence. A double-stranded SMI sequence includes a first degenerate or semi-degenerate nucleotide n-mer sequence and a second n-mer sequence that is complementary to the first degenerate or semi-degenerate nucleotide n-mer sequence, while a single-stranded SMI sequence includes a first degenerate or semi-degenerate nucleotide n-mer sequence. The first and/or second degenerate or semi-degenerate nucleotide n-mer sequences may be any suitable length to produce a sufficiently large number of unique tags to label a set of sheared DNA fragments from a segment of DNA. Each n-mer sequence may be between approximately 3 to 20 nucleotides in length. Therefore, each n-mer sequence may be approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length. In one embodiment, the SMI sequence is a random degenerate nucleotide n-mer sequence which is 12 nucleotides in length. A 12 nucleotide SMI n-mer sequence that is ligated to each end of a target nucleic acid molecule, as described in the Example below, results in generation of up to 4^{24} (i.e., 2.8×10^{14}) distinct tag sequences.

[0027] In some embodiments, the SMI tag nucleotide sequence may be completely random and degenerate, wherein each sequence position may be any nucleotide. (i.e., each position, represented by "X," is not limited, and may be an adenine (A), cytosine (C), guanine (G), thymine (T), or uracil (U)) or any other natural or non-natural DNA or RNA nucleotide or nucleotide-like substance or analog with base-pairing properties (e.g., xanthosine, inosine, hypoxanthine, xanthine, 7-methylguanine, 7-methylguanosine, 5,6-dihydrouracil, 5-methylcytosine, dihydouridine, isocytosine, isoguanine, deoxynucleosides, nucleosides, peptide nucleic acids, locked nucleic acids, glycol nucleic acids and threose nucleic acids). The term "nucleotide" as described herein, refers to any and all nucleotide or any suitable natural or non-natural DNA or RNA nucleotide or nucleotide-like substance or analog with base pairing properties as described above. In other embodiments, the sequences need not contain all possible bases at each position. The degenerate or semi-degenerate n-mer sequences may be generated by a polymerase-mediated method described in the Example below, or may be generated by preparing and annealing a library of individual

oligonucleotides of known sequence. Alternatively, any degenerate or semi-degenerate n-mer sequences may be a randomly or non-randomly fragmented double stranded DNA molecule from any alternative source that differs from the target DNA source. In some embodiments, the alternative source is a genome or plasmid derived from bacteria, an organism other than that of the target DNA, or a combination of such alternative organisms or sources. The random or non-random fragmented DNA may be introduced into SMI adaptors to serve as variable tags. This may be accomplished through enzymatic ligation or any other method known in the art.

[0028] In some embodiments, the SMI adaptor molecules are ligated to both ends of a target nucleic acid molecule, and then this complex is used according to the methods described below. In certain embodiments, it is not necessary to include n-mers on both adapter ends, however, it is more convenient because it means that one does not have to use two different types of adaptors and then select for ligated fragments that have one of each type rather than two of one type. The ability to determine which strand is which is still possible in the situation wherein only one of the two adaptors has a double-stranded SMI sequence.

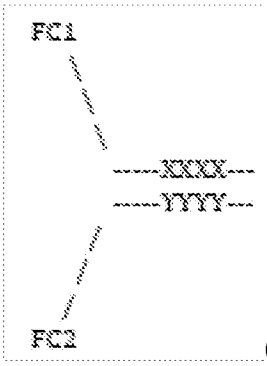
[0029] In some embodiments, the SMI adaptor molecule may optionally include a double-stranded fixed reference sequence downstream of the n-mer sequences to help make ligation more uniform and help computationally filter out errors due to ligation problems with improperly synthesized adaptors. Each strand of the double-stranded fixed reference sequence may be 4 or 5 nucleotides in length sequence, however, the fixed reference sequence may be any suitable length including, but not limited to 3, 4, 5 or 6 nucleotides in length.

[0030] The SMI ligation adaptor may be any suitable ligation adaptor that is complementary to a ligation adaptor added to a double-stranded target nucleic acid sequence including, but not limited to a T-overhang, an A-overhang, a CG overhang, a blunt end, or any other ligatable sequence. In some embodiments, the SMI ligation adaptor may be made using a method for A-tailing or T-tailing with polymerase extension; creating an overhang with a different enzyme; using a restriction enzyme to create a single or multiple nucleotide overhang, or any other method known in the art.

[0031] According to the embodiments described herein, the SMI adaptor molecule may include at least two PCR primer or "flow cell" binding sites: a forward PCR primer binding site (or a "flow cell 1" (FC1) binding site); and a reverse PCR primer binding site (or a "flow cell 2" (FC2) binding site). The SMI adaptor molecule may also include at least two sequencing primer binding sites, each corresponding to a sequencing read. Alternatively, the sequencing primer binding sites may be added in a separate step by inclusion of the necessary sequences as tails to the PCR primers, or by ligation of the needed sequences. Therefore, if a double-stranded target nucleic acid molecule has an SMI adaptor molecule ligated to each end, each sequenced strand will have two reads - a forward and a reverse read.

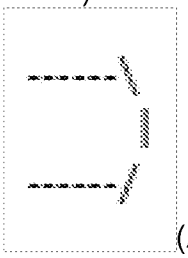
Double-stranded SMI sequences

[0032] Adaptor 1 (shown below) is a Y-shaped SMI adaptor as described above (the SMI sequence is shown as X's in the top strand (a 4-mer), with the complementary bottom strand sequence shown as Y's):

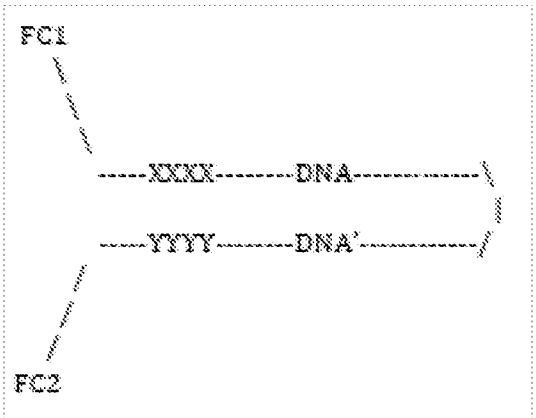


(Adaptor 1)

[0033] Adaptor 2 (shown below) is a hairpin (or "U-shaped") linker:


(Adaptor 2)

[0034] Following ligation of both adaptors to a double-stranded target nucleic acid, the following is structure is obtained:



[0035] When melted, the product will be of the following form (where "linker" is the sequence of adaptor 2):

FC1-------XXXX------DNA---------linker---------DNA'----------YYYY--------FC2

[0036] This product is then PCR amplified. The reads will yield:

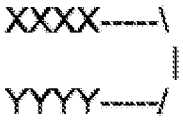
Read 1:
XXXX-----DNA----

Read 2 (note that read 2 is seen as the complement of the bases sequenced:)
XXXX-----DNA----

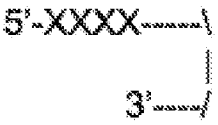
[0037] The sequences of the two duplex strands seen in the two sequence reads may then be compared, and sequence information and mutations will be scored only if the sequence at a given position matches in both of the reads.

[0038] This approach does not strictly require the use of an SMI tag, as the sheared ends can be used as identifiers to differentiate unique individual molecules from PCR duplicates. Thus the same concept would apply if one used any standard sequencing adaptor as "Adaptor 1" and the U-- - shaped linker as "Adaptor 2." However described below, there are a limited number of shear points flanking any given genomic position and thus the power to sequence deeply is increased via inclusion of the SMI tag. A hybrid method using a combination of sheared ends and a shorter n-mer tag (such as 1 or 2 or 3 or 4 or more degenerate or semi-degenerate bases) in the adaptor may also serve as unique molecular identifiers. Another design may include use of any sequencing adaptor (such as one lacking an n-mer tag) in conjunction with an n-mer tag that is incorporated into the U-shaped linker molecule. Such a design would be of the

following form (where X and Y represent complementary degenerate or semi-degenerate nucleotides):

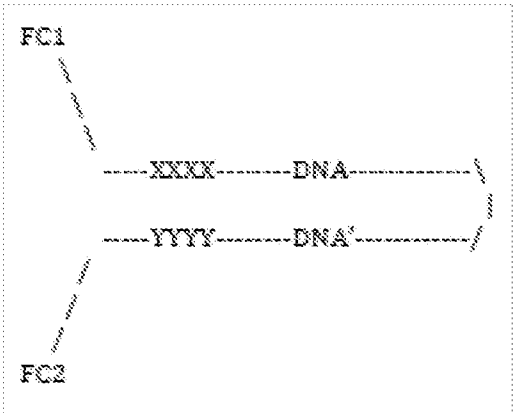


[0039] Synthesis of such a design may be obtained in a number of ways, for example synthesizing a set of hairpin oligonucleotides in which each individual oligonucleotide encodes a complementary n-mer sequence, or alternatively by using a DNA polymerase to carry out extension from the following product (where X's represent degenerate nucleotides):



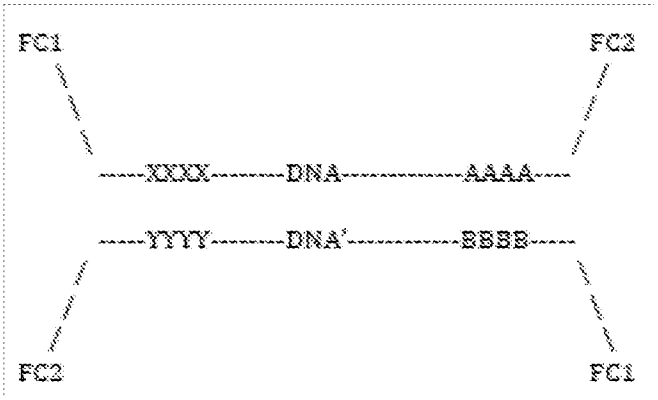
[0040] Inclusion of the SMI tag is also extremely useful for identifying correct ligation products, as the assay uses two distinct adaptors. This will yield multiple possible ligation products:

[0041] **Product I.** Adaptor 1-- - -- - -- - DNA-- - -- - -- - Adaptor 2, which yields the desired product:

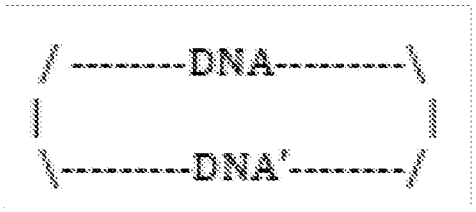


[0042] **Product II.** Adaptor 1-- - -- - -- - DNA-- - -- - -- - Adaptor 1. This will result in the DNA being amplified as two separate strands, i.e. as occurs in the DCS approach described elsewhere in this document (the second copy of Adaptor 1 is shown

below with the SMI as AAA – BBB to emphasize that every DCS adaptor has a distinct SMI sequence)



[0043] **Product III.** Adaptor 2-- - -- - -- - DNA-- - -- - -- - Adaptor 2. This will result in a non-- - amplifiable circular product shown below:



[0044] Product III is non - amplifiable, given the absence of primer binding sites and thus will not be present in the final DNA sequences. Thus only Product II needs to be avoided. The formation of Product II can be minimized in the ligation step by using an excess of Adaptor 2 (relative to Adaptor 1). Then primarily Products I and III will be obtained, with minimal formation of Product II. Additionally, a variety of biochemical means of enriching for products containing adaptor 2 are possible such as using affinity probes that are complementary to the hairpin loop sequence itself. Product I results in the same SMI sequence in both the Read 1 and Read 2 sequence reads. In the example depicted above, Product I sequences can thus be identified by virtue of having matching SMIs of the form XXXX in Read 1 and XXXX in Read 2.

[0045] By contrast, in the case of Product II, the SMI sequences on either end of the sequenced molecule will arise from distinct DCS adaptors having different SMI sequences. In the example shown above, Product II sequences yield SMIs of the form

XXXX (Read 1) – BBB (Read 2) upon sequencing of the top strand, and BBBB (Read 1) – XXXX (Read 2) upon sequencing of the bottom strand. Thus Product II sequences can be easily identified and computationally removed from the final sequence data.

[0046] Data resulting from Product II is useful, because Product II corresponds to the product analyzed under the approach detailed in the Example below. Product I contains a self--complementary hairpin sequence that can impair polymerase extension during amplification, however, this type of amplification has already been enabled in the technique of "Hairpin PCR" [50] which involves linking of the two strands followed by amplification with gene-specific primers. Amplification conditions that are compatible with amplification of hairpin DNA are thus already established. Moreover, ligation and amplification with circularizing "linkers" (i.e. hairpin linkers affixed to both ends of a fragment) has been demonstrated as a step in the Pacific Biosciences sample preparation workflow [49]. As the sequence of the linker itself does not matter in the workflow, the published linker sequences from either of these references would be adequate for use in the assay.

[0047] In some aspects of some embodiments, deliberate ligation of "U-shaped" adaptors or hairpin linkers containing 1) a double-stranded n-mer (or other form of degenerate or semi-degenerate double-stranded tag as enumerated above) plus 2) primer binding sites to both ends of a captured fragment may be desireable. Producing closed circles of captured material may help facilitate removal of non-captured DNA by exonuclease digestion given that circularized DNA will be protected from digestion by such enzymes. Additionally, closed circles may be pre-amplified using rolling circle amplification or serve as the substrate for continuous loop sequencing [49]. Recognition sites for restriction endonuclease digestion could be engineered into these adaptors to render closed loops open once again if more convenient for subsequent steps.

[0048] In another embodiment, flow cell sequences or PCR binding sites, again denoted as FC1 and FC2, may be included in both the PCR primers and the hairpin linker adaptor, as well as a ligatable sequence on the end of the hairpin linker (denoted as L below). The hairpin linker adaptor may additionally include one or more cleavable

sequences, denoted as R in the example below (the R may be any appropriate restriction enzyme target sequence, or any other cleavable sequence). Such a hairpin linker design is shown below:

```
L-XXXX--FC2-R----\
                  |
L-YYYY--FC1-R----/
```

[0049] The target DNA with ligation site denoted as L is as follows:

```
--------------DNA--------L
--------------DNA'-------L
```

[0050] Following ligation of the linker, the product may be amplified with PCR primers as follows:

```
FC1
   \
    \
     \
      ------
      --------------DNA-------- XXXX--FC2-R----\
                                                |
      --------------DNA'------- YYYY--FC1-R----/
            ------
          /
         /
        /
      FC2
```

[0051] The resultant product will be of the form:

```
FC1--------DNA--------XXXX--FC2-R-----R-FC1--YYYY--------DNA'------- FC2
```

[0052] After amplification of the product, the cleavage sites R may be cleaved to result in the following sequencable products:

```
FC1--------DNA-------- XXXX--FC2
```

and

```
FC1--YYYY--------DNA'------- FC2
```

[0053] These products may then be sequenced directly. This design has the advantage of allowing for targeted sequencing of a specific region of the genome, and furthermore avoids the need to sequence a hairpin product, as sequencing of a hairpin will be less efficient due to the self-complementarity present within the hairpin molecule.

Single-stranded SMI sequences

[0054] In one embodiment, a single-stranded SMI sequence is incorporated into the single-stranded portion of the hairpin loop (regions of sequence complementarity are denoted as "="). The SMI sequence is shown as four nucleotides in length in the following examples, but in practice an Nmer of any length, including approximately 3 to 20 nucleotides, will suffice.

```
=====NNNN-----\
              |
=====----------/
```

[0055] Ligation of the hairpin linker and a Y-shaped sequencing adaptor (with PCR primer binding sites labeled as FC1 and FC2) yields the following product:

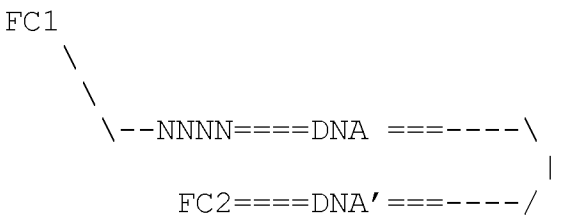
```
 FC1
    \
     \
      =====DNA =====NNNN------\
                             |
      =====DNA'==============/
     /
    /
   /
FC2
```

[0056] Melting and PCR amplification of this product yields the following DNA product:

```
FC1-------DNA-----NNNN------hairpin sequence------DNA'-------FC2
```

[0057] Following PCR duplication of the product and formation of consensus reads based upon the shared SMI sequence among all the PCR duplicates, the sequences of the two strands (denoted DNA and DNA') can then be compared to form a duplex consensus sequence.

[0058] In another embodiment, a single-stranded SMI is incorporated into a modified "Y-shaped" sequencing adaptor in which PCR primer binding sites are located at the sites labeled FC1 and FC2 (regions of sequence complementarity are depicted as "=")



[0059] It will be apparent to one skilled in the art that a single-stranded SMI sequence tag can be located in any of several positions within either the sequencing adaptor or the hairpin linker. The single-stranded SMI sequence tag can be synthesized as a random oligonucleotide sequence, or can be sequenced as a set of fixed sequences by synthesis on an array, or by any other suitable method known in the art.

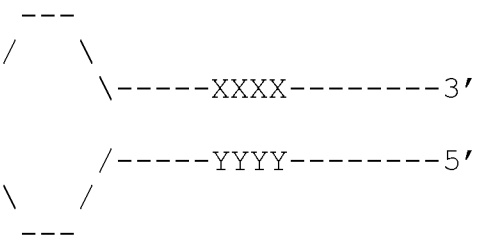
Methods for synthesis of complementary or partially complementary double stranded SMI tags

[0060] SMI adaptors molecules containing a double-stranded, complementary, degenerate or semi-degenerate SMI tag can be made by any of a number of methods, including copying of a single-stranded SMI sequence by a DNA polymerase as described above or synthesis and annealing of two oligonucleotides containing complementary SMI sequences. An additional method involves synthesizing a set of linear oligonucleotides which will self-anneal into the appropriate form. Inclusion of a cleavable linker in each oligonucleotide will then allow for conversion of a "hairpin shaped" SMI adaptor molecule into a "Y-shaped" SMI adaptor molecule. For example, an oligonucleotide may be prepared of the following form:

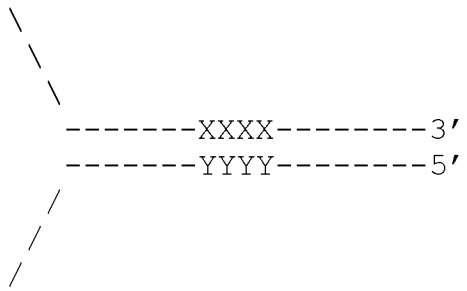
5'---------YYYY-------U-------XXXX----------3'

[0061] In this schematic, X and Y represent complementary nucleotides, and U indicates a cleavable linker, such as uracil (which can be cleaved by combined

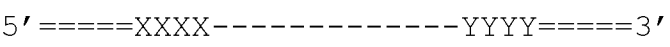
treatment with uracil DNA glycosylase and apurinic endonuclease), although any other cleavable linker will suffice. The oligonucleotide may be designed with appropriate regions of self-complementarity to anneal into the following form:



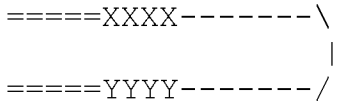
[0062] The linker (e.g. uracil) may then be cleaved, yielding a DCS adaptor:



[0063] A double-stranded SMI hairpin linker can be constructed by an analogous method but without the need for a cleavable linker. For example, a set of nucleotides of known sequence where X and Y represent the complementary SMI sequences can be synthesized on an array, or by any other suitable method known in the art:



[0064] This oligonucleotide can then self-anneal to form a hairpin linker with complementary SMI sequences.



[0065] Any of the oligonucleotides described above can also include any ligatable sequence as overhangs on either the 5' or 3' end, or can be used for blunt end ligation.

DCS SMI adaptor molecules may include sequences to allow for targeted DNA capture

[0066] DCS SMI adaptor molecules contain ligatable ends to allow attachment of the adaptor to a target DNA molecule. In some embodiments, the ligatable end may be complementary to a DNA overhang on the target DNA, for example, one generated by digestion of target DNA with a restriction endonuclease. Selective ligation of the adaptor to the targeted DNA containing the matching Single-stranded overhanging DNA sequence will then allow for partial purification of the targeted DNA. A non-limiting example of this embodiment is shown above ~~in paragraphs [0048]-[0053]~~. In some embodiments, the DCS SMI adaptor molecule, or a hairpin linker SMI adaptor molecule, may additionally contain modifications such as biotin to facilitate affinity purification of target DNA that has ligated to the adaptor.

[0067] In another embodiments, specific PCR primers can selectively amplify specific regions of genome when the adaptor that is ligated to the other end of the molecule is a hairpin (or "U-shape"). Alternatively, this method may be used with or without the need for this cleavable hairpin sequence.

Preparation of DNA for duplex consensus sequencing may be performed by PCR amplification in a hairpin structure

[0068] Another embodiment involves fragmentation of DNA at defined regions, for example by treatment of DNA with a site-specific restriction endonuclease or a mixture of such endonucleases, followed by annealing of a hairpin oligonucleotide linker, and amplification of the hairpin complex with PCR primers sufficient for amplification of the desired DNA sequence. Annealing of the hairpin linker to only one of the two ends of the DNA duplex could be accomplished by using different restriction enzymes to cut on either end of the target duplex, and then having the hairpin linker ligation adaptor being ligatable to only one of the two resultant ligatable ends.

[0069] The example shown below indicates forward and reverse PCR primers (labeled 1 and 2) in conjunction with a hairpin linker to allow linked amplification of both complementary strands of duplex DNA. Such amplification, in conjunction with a single-stranded or double-stranded SMI sequence, would allow for targeted amplification and high accuracy deep sequencing of a specific sequence of interest. In the schematic shown below, a single-stranded SMI sequence is incorporated into PCR primer FC1. It would be apparent to one skilled in the art that the SMI sequence could also be incorporated in primer FC2, or in the hairpin linker.

```
FC1NNNN___
         DNA ==========\
                       |
         DNA'==========/
      FC2___
```

[0070] Amplified product:

```
FC1NNNN DNA----hairpin sequence----DNA'FC2
```

[0071] This product can then be subjected to consensus sequencing analysis. The SMI sequence allows one to group together products of PCR amplification arising from a single molecule of duplex DNA. The sequences of the two DNA strands can then be compared for error correction.

Uses of SMI adapter molecules

[0072] The SMI adaptor molecules described herein have several uses. In some embodiments, the SMI adaptor molecules described herein may be used in methods to obtain the sequence or other sequence-related information of a double-stranded target nucleic acid molecule. According to the embodiments described herein, the term "double-stranded target nucleic acid molecule" includes a double-stranded DNA molecule or a double-stranded RNA molecule. Thus, the SMI adaptor molecules and methods of use described herein are applicable to genotyping and other applications related to sequencing of DNA molecules, but are also applicable to RNA sequencing applications such as for sequencing of double-stranded RNA viruses. Methods for sequencing RNA may include any of the embodiments described herein with respect to

DNA sequencing, and vice-versa. For example, any double stranded target nucleic acid molecule may be ligated to an SMI adaptor molecule which includes a double-stranded RNA or DNA n-mer tag and an RNA or DNA ligation adapter as described above. Methods exist for directly sequencing RNA [51]; alternatively, the ligated product may be reverse transcribed into DNA, and then sequenced as a double-stranded target DNA molecule.

[0073] In one embodiment, the double-stranded target nucleic acid molecule may be a sheared double-stranded DNA or RNA fragment. The sheared target DNA or RNA molecule may be end repaired and a double-stranded target nucleic acid sequence ligation adaptor may be added to each end of the sheared target DNA or RNA molecule. The double-stranded target nucleic acid sequence ligation adaptor may be any suitable ligation adaptor that is complementary to the SMI ligation adaptor described above including, but not limited to a T-overhang, an A-overhang, a CG overhang, blunt end or any other ligatable sequence. In some embodiments, the double-stranded target nucleic acid sequence ligation adaptor may be made using a method for A-tailing or T-tailing with polymerase extension; adding an overhang with a different enzyme; using a restriction enzyme to create a ligatable overhang; or any other method known in the art.

[0074] Methods to obtain the sequence or other sequence-related information of a double-stranded target nucleic acid molecule may include a step of ligating the double-stranded target nucleic acid molecule to at least one SMI adaptor molecule, such as those described above, to form a double-stranded target nucleic acid complex. In one embodiment, each end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule. The double-stranded target nucleic acid complex is then amplified by a method known in the art (e.g., a PCR or non-PCR method known in the art), resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products. These products are then sequenced using any suitable method known in the art including, but not limited to, the Illumina sequencing platform, ABI SOliD sequencing platform, Pacific Biosciences sequencing platform, 454 Life Sciences sequencing platform, Ion Torrent sequencing platform, Helicos sequencing platform, and nanopore sequencing technology.

[0075] In certain embodiments, a method of generating an error corrected double-stranded consensus sequence is provided. Such a method, also referred to as duplex consensus sequencing (DCS), allows for a quantitative detection of sites of DNA damage. DCS analysis facilitates the detection of DNA damage signatures, in that single stranded DNA mutations that are not present in the complementary strand can be inferred to be artifactual mutations arising from damaged nucleotides. Not only can one correct for these erroneous mutations, but the ability to indirectly infer that damage is present on the DNA could be a useful biomarker (e.g. for cancer risk, cancer metabolic state, mutator phenotype related to defective damage repair, carcinogen exposure, chronic inflammation exposure, individual-specific aging, neurodegenerative diseases etc). The ability to use different polymerases during the first round(s) of PCR to mis-incorporate at damage sites could potentially add even more information. Besides polymerases, other DNA modifying/repair enzymes could be used prior to amplification to convert damage of one sort that doesn't give a specific mutagenic signature into another sort that does with whatever polymerase is used. Alternatively, DNA modifying/repair enzymes could be used to remove damaged bases, and one could sequence both strands of DNA both with and without the enzymatic treatment. Mutations in single-stranded DNA that are seen to be removed by the enzymatic treatment can thus be inferred to be arising due to DNA damage. This could be useful on human nuclear or mtDNA but also might also be useful with model organisms (mice, yeast, bacteria etc), treated with different new damaging agents, facilitating a screen for DNA damaging compounds that would be analogous to the widely used Ames test [52].

[0076] The method of generating an error corrected double-stranded consensus sequence may include a first stage termed “single strand consensus sequencing” (SSCS) followed by a second stage of duplex consensus sequencing (DCS). Therefore, the method includes steps of tagging individual duplex DNA molecules with an SMI adaptor molecule, such as those described above; generating a set of PCR duplicates of the tagged DNA molecules by performing a suitable PCR method; creating a single strand consensus sequence from all of the PCR duplicates which arose from an individual molecule of single-stranded DNA. Each DNA duplex should result in two

single strand consensus sequences. The work through these three steps conclude the first stage and is termed SSCS.

[0077] The method of generating an error corrected double-stranded consensus sequence further comprises the second stance that is termed DCS. The DCS stage includes steps of comparing the sequence of the two single strand consensus sequences arising from a single duplex DNA molecule, and further reducing sequencing or PCR errors by considering only sites at which the sequences of both single-stranded DNA molecules are in agreement. The method that includes the first stage and the second stage termed Duplex Consensus Sequencing (DCS).

[0078] The step of tagging of both strands of individual duplex DNA may be accomplished by ligation of degenerate or semi-degenerate complementary DNA sequences; as the complementary nature of the two strands of such a tag sequence allows the two molecules to be grouped together for error correction. Alternatively, as described above, the two duplex DNA strands may be linked by ligation of a U-shaped SMI adaptor molecule, and the two DNA strands can thus both be tagged with a single-stranded SMI tag.

[0079] In the method described above, a set of sequenced SMI-DNA products generated in the methods described above may be grouped into families of paired target nucleic acid strands based on a common set of SMI sequences. Then, the paired target nucleic acid strands can be filtered to remove nucleotide positions where the sequences seen on both of the paired partner DNA strands are not complementary. This error corrected double-stranded consensus sequence may be used in a method for confirming the presence of a true mutation (as opposed to a PCR error or other artifactual mutation) in a target nucleic acid sequence. According to certain embodiments, such a method may include identifying one or more mutations present in the paired target nucleic acid strands that have one or more nucleotide positions that disagree between the two strands, then comparing the mutation present in the paired target nucleic acid strands to the error corrected double-stranded consensus sequence. The presence of a true mutation is confirmed when the mutation is present on both of

the target nucleic acid strands and also appear in all members of a pared target nucleic acid family.

[0080] The accuracy of current approaches to next-generation sequencing is limited due to their dependence on interrogating single-stranded DNA. This dependence makes potential sources of error such as PCR amplification errors and DNA damage fundamentally limiting. However, the complementary strands of a double-stranded DNA molecule (or "DNA duplex") contain redundant sequencing information (i.e., one molecule reciprocally encoding the sequence information of its partner) which can be utilized to eliminate such artifacts. Limitations related to sequencing single-stranded DNA (e.g., sequencing errors) may therefore be overcome using the methods described herein. This is accomplished by individually tagging and sequencing each of the two strands of a double-stranded (or duplex) target nucleic acid molecule and comparing the individual tagged amplicons derived from one half of a double-stranded complex with those of the other half of the same molecule. Duplex Consensus Sequencing (DCS), significantly lowers the error rate of sequencing. In some embodiments, the DCS method may be used in methods for high sensitivity detection of rare mutant and variant DNA as described further below.

[0081] As described above, one approach that has previously been reported for DNA sequencing involves incorporation of a random tag sequence into a PCR primer [36]. This approach results in an improvement in accuracy relative to standard Illumina sequencing, but is fundamentally limited in that it is based upon amplification and sequencing of single-stranded DNA and thus cannot overcome limitations in sensitivity owing to single-stranded DNA damage events. In the methods described herein, PCR duplicates are generated from a single strand of DNA, and the sequences of the duplicates are compared. Mutations are scored only when they are present in multiple replicates of a single starting molecule. The DCS approach overcomes the limitation of previous approaches by considering both DNA strands.

[0082] DNA damage should not be a limiting factor in DCS, because miscoding damage events at a single base-pair position occur essentially exclusively on only one of the two DNA strands. For DNA damage to result in an artifactual mutation in DCS,

damage would need to be present at the same nucleotide position on both strands. Even if complementary nucleotides in a duplex were both damaged, the damage would need to result in complementary sequencing errors to result in mis-scoring of a mutation. Likewise, spontaneous PCR errors would need to result in complementary mutations at the same position on both strands; with a first-round mutation frequency of Taq polymerase of approximately 10^{-5} and three possible incorrect bases that could be mis-inserted, the probability of two complementary PCR errors occurring would be $10^{-5} \times 10^{-5} \times 1/3 = 3.3 \times 10^{-11}$

[0083] According to some embodiments, the sequencing method may be performed using the Illumina or similar platforms including those enumerated above without the use of SMI adaptor molecules, but instead by using the random shear points of DNA as identifiers. For a given DNA sequence seen in sequencing read 1 with a specific set of shear points, the partner strand will be seen as a matching sequence in read two with identical shear points. In practice, this approach is limited by the limited number of possible shear points that overlap any given DNA position. However, according to some embodiments, shear points of a target nucleic acid molecule may be used as unique identifiers to identify double-stranded (or duplex) pairs, resulting in an apparent error frequency at least as low as that seen with traditional sequencing methods, but with a significantly lower loss of sequence capacity. In other embodiments, DCS based on shear points alone may have a role for confirmation that specific mutations of interest are true mutations which were indeed present in the starting sample (i.e. present in both DNA strands), as opposed to being PCR or sequencing artifacts. Overall, however, DCS is most generally applicable when randomized, complementary double-stranded SMI sequences are used. A 24 nucleotide double-stranded SMI sequence was used in the Example described below, which may yield up to $4^{24} = 2.8 \times 10^{14}$ distinct double-stranded SMI sequences. Combining information regarding the shear points of DNA with the SMI tag sequence would allow a shorter SMI to be used, thus minimizing loss of sequencing capacity due to sequencing of the SMI itself.

[0084] In certain embodiments, the SMI adaptor molecules may also be used in methods of single-molecule counting for accurate determination of DNA or RNA copy

number [38]. Again, since the SMI tags are present in the adaptors, there are no altered steps required in library preparation, which is in contrast to other methods for using random tags for single-molecule counting. Single-molecule counting has a large number of applications including, but not limited to, accurate detection of altered genomic copy number (e.g., for sensitive diagnosis of genetic conditions such as trisomy 21 [47]), for accurate identification of altered mRNA copy number in transcriptional sequencing and chromatin immunoprecipitation experiments, quantification of circulating microRNAs, quantification of viral load of DNA or RNA viruses, quantification of microorganism abundance, quantification of circulating neoplastic cells, counting of DNA-labeled molecules of any variety including tagged antibodies or aptamers, and quantification of relative abundances of different individual's genomes in forensic applications.

[0085] In another embodiment, the SMI adaptor molecules may be used in methods for unambiguous identification of PCR duplicates. In order to restrict sequencing analysis to uniquely sequenced DNA fragments, many sequencing studies include a step to filter out PCR duplicates by using the shear points at the ends of DNA molecules to identify distinct molecules. When multiple molecules exhibit identical shear points, all but one of the molecules are discarded from analysis under the assumption that the molecules represent multiple PCR copies of the same starting molecule. However sequence reads with identical shear points can also reflect distinct molecules because there are a limited number of possible shear points at any given genomic location, and with increasing sequencing depth, recurrent shear points are increasingly likely to be seen [48]. Because the use of SMI tags (or "double-stranded SMI sequences") allows every molecule to be uniquely labeled prior to PCR duplication, true PCR duplicates may be unambiguously identified by virtue of having a common (i.e., the same or identical) SMI sequence. This approach would thereby minimize the loss of data by overcoming the intrinsic limitations of using shear points to identify PCR duplicates.

[0086] Importantly, once SMI-containing adaptors are synthesized by a straightforward series of enzymatic steps or are produced through synthesis of a set of oligonucleotides containing complementary tag sequences, they may be substituted for

standard sequencing adaptors. Thus, use of DCS does not require any significant deviations from the normal workflow of sample preparation for Illumina DNA sequencing. Moreover, the DCS approach can be generalized to nearly any sequencing platform because a double-stranded SMI tag can be incorporated into other existing adaptors, or for sequencing approaches that do not require adaptors, a double-stranded SMI tag can be ligated onto duplex DNA sample prior to sequencing. The compatibility of DCS with existing sequencing workflows, the potential for greatly reducing the error rate of DNA sequencing, and the multitude of applications for the double-stranded SMI sequences validate DCS as a technique that may play a general role in next generation DNA sequencing.

[0087] The following examples are intended to illustrate various embodiments of the invention. As such, the specific embodiments discussed are not to be construed as limitations on the scope of the invention. It will be apparent to one skilled in the art that various equivalents, changes, and modifications may be made without departing from the scope of invention, and it is understood that such equivalent embodiments are to be included herein. Further, all references cited in the disclosure are hereby incorporated by reference in their entirety, as if fully set forth herein.

EXAMPLES

Example 1: Generation of SMI Adaptor Molecules and their use in sequencing double-stranded target DNA

Materials and Methods

[0088] *Materials.* Oligonucleotides were from IDT and were ordered as PAGE purified. Klenow exo- was from NEB. T4 ligase was from Enzymatics.

[0089] *DNA isolation.* Genomic DNA was isolated from normal human colonic mucosa by sodium iodide extraction (Wako Chemicals USA).

[0090] *Adaptor synthesis.* The adaptors were synthesized from two oligos, designated as:

the primer strand:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT (SEQ ID NO:1); and

the template strand:

/5phos/ACTGNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTG AACTCCAGTCAC (SEQ ID NO:2).

[0091] The two adaptor strands were annealed by combining equimolar amounts of each oligo to a final concentration of 50 micromolar and heating to 95°C for 5 minutes. The oligo mix was allowed to cool to room temperature for over 1 hour. The annealed primer-template complex was extended in a reaction consisting of 40 micromolar primer-template, 25 units Klenow exo-DNA polymerase (New England Biolabs), 250 micromolar each dNTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, and 1 mM dithiothreitol (DTT) for 1 hour at 37°C. The product was isolated by ethanol precipitation. Due to the partial A-tailing property of Klenow exo-, this protocol results in a mixture of blunt-ended adapters and adapters with a single-nucleotide A hverhang. A single-nucleotide A overhang was added to residual blunt fragments by incubating the adapters with 25 units Klenow exo-, 1 mM dATP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM MgCl2, and 1 mM dithiothreitol (DTT) for 1 hour at 37°C. The product was again ethanol precipitated and resuspended to a final concentration of 50 micromolar.

[0092] *Sequencing library preparation.* 3 micrograms of DNA was diluted into 130 microliters of TE buffer (10 mM tris-HCl, pH 8.0, 0.1 M EDTA) and was sheared on the Covaris AFA system with duty cycle 10%, intensity 5, cycles/burst 200, time 20 seconds x 6, temperature 4°C. DNA was purified with 2 volumes of Agencourt AMPure XP beads per the manufacturer's protocol. After end-repair with the NEB end-repair kit per the manufacturer's protocol, DNA fragments larger than the optimal range of ~200-500 bp were removed by adding 0.7 volumes of AMPure XP beads and transferring the supernatant to a separate tube (fragments larger than 500 bp bind to the beads and are discarded). An additional 0.65 volumes of AMPure XP beads were added (this step allows fragments of approximately 200bp or greater to bind to the

beads). The beads were washed and DNA eluted. DNA was then T-tailed in a reaction containing 5 units Klenow exo-, 1 mM dTTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM MgCl2, 1 mM. The reaction proceeded for 1 hour at 37 C. DNA was purified with 1.2 volumes of AMPure XP beads. The custom adaptors were ligated by combining 750 ng of T-tailed DNA with 250 pmol adaptors in a reaction containing 3000 units T4 DNA ligase, 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 5 mM DTT, 1 mM ATP. The reaction was incubated 25 C for 15 minutes, and purified with 1.2 volumes of AMPure XP beads.

[0093] *Pre-capture amplification.* 375 ng adaptor-ligated DNA was PCR amplified with primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and GTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:4) using the Kappa high-fidelity PCR kit for 8 cycles with an annealing temperature of 60 C. The product was purified with 1.2 volumes of AMPure XP beads.

[0094] *DNA capture.* Target capture was performed with the Agilent SureSelect system per the manufacturer's recommendations, except that capture volumes were performed at one-half of the standard volume. The capture set targeted an arbitrary 758 kb region of the genome consisting of both coding and noncoding sequences. Capture baits were 120 nt in length, and were prepared with the Agilent eArray tool with 3x tiling.

[0095] *Post-capture amplification.* Captured DNA was amplified with PCR primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and CAAGC AGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:5) where XXXXXX indicates the position of a fixed multiplexing barcode sequence). 2 0 fmol of DNA was used per lane for sequencing on an Illumina HiSeq 2000.

[0096] *Data processing.* Reads with intact SMI adaptors include a 12 nucleotide random sequence, followed by a 5 nucleotide fixed sequence. These reads were identified by filtering out reads that lack the expected fixed sequence at positions 13-17. The SMI sequence from both the forward and reverse sequencing reads (i.e., the first and second degenerate n-mer sequences) was computationally added to the read header, and the fixed sequence removed. The first 4 nucleotides located following the

adaptor sequence were also removed due to the propensity for ligation and end-repair errors to result in an elevated error rate near the end of the DNA fragments. Reads having common (i.e., identical) SMI sequences were grouped together, and were collapsed to generate a consensus read. Sequencing positions were discounted if the consensus group covering that position consisted of fewer than 3 members, or if fewer than 90% of the sequences at that position in the consensus group had the identical sequence. Reads were aligned to the human genome with the Burrows-Wheeler Aligner (BWA). The consensus sequences were then paired with their strand-mate by grouping each 24 nucleotide tag of form AB in read 1 with its corresponding tag of form BA in read 2. Resultant sequence positions were considered only when information from both DNA strands was in perfect agreement. An overview of the data processing workflow is as follows:

1. Discard reads that do not have the 5 nt fixed reference (or "spacer") sequence (CAGTA; SEQ ID NO:6) present after 12 random nucleotides.
2. Combine the 12 nt SMI tags from read 1 and read 2, and transfer the combined 24 nt SMI sequence into the read header.
3. Discard SMIs with inadequate complexity (i.e., those with > 10 consecutive identical nucleotides).
4. Remove the 5 nt fixed reference sequence.
5. Trim an additional 4 nt from the 5' ends of each read pair (sites of error prone end repair).
6. Group together reads which have identical 24 nt SMIs.
7. Collapse to SMI consensus reads, scoring only positions with 3 or more SMI duplicates and >90% sequence identity among the duplicates.
8. For each read in read 1 file having SMI of format AB, group with corresponding DCS partner in read 2 with SMI of format BA.
9. Only score positions with identical sequence among both DCS partners.
10. Align reads to the human genome.

[0097] Code for carrying out the workflow may be pre-existing or may involve programming within the skill of those in the art. In some embodiments, however, the

Python code, which is illustrated in Figure 10, may be used for carrying out the pairing and scoring of partner strands according to steps 8 and 9 of the workflow described above.

Overview

[0098] To overcome limitations in the sensitivity of variant detection by single-stranded next-generation DNA sequencing, an alternative approach to library preparation and analysis was designed, which is known herein as Duplex Consensus Sequencing (DCS) (Figure 1). The DCS method described herein involves tagging both strands of duplex DNA with a random, yet complementary double-stranded nucleotide sequence, which is known herein as a double-stranded single molecule identifier (SMI) sequence. The SMI sequences (in this case, double stranded SMI sequences) are incorporated into the SMI adaptor molecules by introducing a single-stranded randomized nucleotide sequence into one adapter strand and the extending the opposite strand with a DNA polysmerase to yield a complementary, double-stranded SMI sequence (Figure 2). The individually tagged strands are then PCR amplified. Every duplicate that arises from a single strand of DNA will have the same SMI, and thus each strand in a DNA duplex pair generates a distinct, yet related population of PCR duplicates after amplification owing to the complementary nature of the SMIs on the two strands of the duplex. Comparing the sequence obtained from each of the two strands comprising a single molecule of duplex DNA facilitates differentiation of sequencing errors from true mutations. When an apparent mutation is, due to a PCR or sequencing error, the substitution will only be seen on a single strand. In contrast, with a true DNA mutation, complementary substitutions will be present on both strands (see Figure 4C).

[0099] Following tagging with a double-stranded SMI and PCR amplification, a family of molecules is obtained that arose from a single DNA molecule; members of the same PCR "family" are then grouped together by virtue of having a common (i.e., the same) SMI tag sequence. The sequences of uniquely tagged PCR duplicates are subsequently compared in order to create a PCR consensus sequence. Only DNA positions that yield the same DNA sequence in a specified proportion of the PCR duplicates in a family, such as 90% of the duplicates in one embodiment, are used to

create the PCR consensus sequence. This step filters out random errors introduced during sequencing or PCR to yield the PCR consensus sequences, each of which derives from an individual molecule of single-stranded DNA. This set of PCR consensus sequences are called single strand consensus sequences (SSCSs).

[00100] Next, PCR consensus sequences arising from two complementary strands of duplex DNA can be identified by virtue of the complementary SMIs (Figure 3) to identify the "partner SMI." Specifically, a 24-nucleotide SMI consists of two 12-nucleotide sequences that can be designated XY. For an SMI of form XY in read 1, the partner SMI will be of form YX in read 2. An example to illustrate this point is given in Figure 4. Following partnering of two strands by virtue of their complementary SMIs, the sequences of the strands are compared. Sequence reads at a given position are kept only if the read data from each of the two paired strands is in agreement.

Results

[00101] In order to label or tag each of the strands of duplex DNA with unique complementary tags, adaptors which contain the standard sequences required for the Illumina HiSeq system were synthesized, but with addition of a double-stranded, complementary SMI sequence (or "tag") of 12 random nucleotides (or a random "degenerate sequence") per strand. Target DNA molecules having a random SMI sequence n-mer that is 12 nucleotides in length on each end will therefore have a unique 24 nucleotide SMI sequence. The adaptors were prepared (Figure 2) from two partially complementary oligonucleotides, one of which has a single-stranded 12 nucleotide random nucleotide sequence (i.e. a first random degenerate nucleotide n-mer sequence) followed by a single stranded fixed reference sequence that is 4 nucleotides in length. The single-stranded random nucleotide tag was converted to a double-stranded, complementary SMI tag by extension with Klenow exo- DNA polymerase and the extended adaptor was purified by ethanol precipitation. Due to the partial A-tailing property of Klenow exo-, this protocol results in a mixture of blunt-ended adaptors and adaptors with a single-nucleotide A overhang (data not shown). A single-nucleotide A-overhang was added to the residual blunt fragments by incubating the adaptors with Klenow exo- DNA polymerase and a high concentration of dATP (1 mM), and purified the adaptors again by ethanol precipitation.

[00102] DNA for sequencing was sheared and end-repaired by standard methods, with size-selection for fragments in the range of ~200-500 bp by size-selective binding to Ampure XP beads. Standard Illumina library preparation protocols involve ligating A-tailed DNA to T-tailed adaptors. However, because A-tailed adaptors were used, the DNA was T-tailed by incubating the end-repaired DNA with Klenow exo- DNA polymerase and 1 mM dTTP. The adaptor-ligated library was PCR amplified and subjected to SureSelect capture, with targeting of an arbitrary 758 kb portion of the genome (DNA coordinates available upon request). The efficiency of adaptor ligation, PCR amplification, DNA capture, and sequencing were comparable to those seen with standard library preparation methods (data not shown). Although Agilent Sure Select probes are used in this example, any suitable method of DNA selection may be used to capture particular target double-stranded DNA sequences. For example, selection and capture may be accomplished by any selection by hybridization method (e.g., Agilent SureSelect, Primer Extension Capture, exploitation of biotinylated PCR amplicons as bait, Agilent HaloPlex) wherein probes that target the desired double-stranded DNA sequence may be recovered by an in-array capture (using probes immobilized on glass slides) or by affinity using magnetic beads in an in-solution capture. In addition, mitochondrial and some other forms of DNA may be isolated by size selection. Alternatively, in some embodiments, no enrichment is performed.

[00103] This protocol was used to sequence DNA isolated from normal colonic mucosa. Mutations were initially scored without consideration of the SMI sequences. PCR duplicates were filtered out with samtools rmdup, a standard tool which uses the shear points of DNA molecules to identify PCR duplicates, as molecules arising from duplicated DNA will have shared shear points. In order to focus specifically on non-clonal mutations, only those positions in the genome with at least 20x coverage and at which fewer than 5% of reads differed from the hg19 reference sequence were considered. This approach resulted in 70.9 million nucleotides of sequence data and 56,890 mutations, indicating an overall mutation frequency of 8.03×10^{-4}, in accord with the error rate of Illumina next-generation sequencing of ~0.1-1% [32].

[00104] Next, the SMI tags were used to group together PCR duplicates that arose from individual single-stranded DNA molecules and to create a consensus sequence

from the family of duplicates. At least 3 PCR duplicates were required, with at least 90% agreement in sequence among all duplicates, to consider a site for mutations. Scoring the mutation frequency as above, again considering only sites with a minimum of 20x coverage and with <5% of reads differing from reference, resulted in 145 million nucleotides of sequence with 6,508 mutations and an overall mutation frequency of $4.47x10^{-5}$, consistent with prior reports [36]. Notably, far more nucleotides of DNA sequence were obtained in this approach (145 million) than in the standard Illumina sequencing approach (70 million) detailed above which is dependent on use of the shear points of single-ended reads to identify PCR duplicates. The improved sequence coverage arose from use of the SMI to identify PCR duplicates, because identifying PCR duplicates by consideration of uniquely sheared DNA ends is fundamentally limited by the small number of possible shear points that overlap a given position of the genome and the propensity for specific genomic regions to be more readily undergo shearing. Thus filtering PCR duplicates by using shear points resulted in discarding a large portion of the reads.

[00105] Finally, the complementary nature of the double-stranded SMI sequences was used to identify pairs of consensus groups that arose from complementary DNA strands. Sequence reads were considered only when the read data from each of the two strands is in perfect agreement. In a pilot experiment, after grouping of PCR duplicates as above, 29,409 SMI partner pairs were found, indicative that fewer than 1% of tags had their corresponding partner tag present in the library. The low recovery of tag pairs was most likely due to inadequate amplification of the starting DNA library. Among these tag-pairs, 24,772 duplex consensus strands were identified with an average strand length of 82 nucleotides, resulting in 2 million nucleotides of DNA consensus sequence. The sequences of the paired duplex strands disagreed at 3,585 of the nucleotide positions, indicative of single-stranded errors (i.e. PCR or sequencing errors); these sites of disagreement were removed, leaving only bases at which the sequence of both duplex strands were in perfect agreement. Next, as above, analysis of mutation frequencies was restricted to sites with at least 10x coverage and at which fewer than 10% of reads disagreed from the hg19 reference sequence. Because the 2 million nucleotides of read data were spread across a 758kb target, our average depth

was only ~3x. Thus only 14,464 nucleotides of DNA sequence corresponded to sites with at least 10x depth. Among these sites, zero mutations were seen. To increase the number of tag pairs considered, analysis described above was repeated, but PCR duplicates were grouped with a minimum of only 1 duplicate per site. This resulted in 28,359 nucleotides of DNA sequence with at least 10x depth. Again, no mutations were detected.

[00106] Current experiments are being performed on vastly smaller target DNA molecules (ranging from ~300 bp to ~20 kb in size). Use of smaller DNA targets will allow for much greater sequencing depth, and far more accurate assessment of the background mutation rate of the assay. In addition, the protocol has been modified to incorporate a greater number of PCR cycles initiated off a smaller number of genome equivalents, which will increase the fraction of tags for which both of the partner tag strands have been sufficiently amplified to be represented in the final sequence data. Indeed, among the 3.6 million SMIs present in our initial library which underwent PCR duplication, 1.5 million of the SMIs were present only once, indicating insufficient amplification of the DNA due in part to the low number of PCR cycles used.

Example 2: Duplex Sequencing of Human Mitochondrial DNA.

Materials and methods

[00107] In addition to those described in Example 1 above, the following materials and methods were also used.

[00108] *DNA Isolation*. Mitochondrial DNA was isolated as previously described (4).

[00109] *Data processing*. The entire human genome sequence (hg19) was used as reference for the mitochondrial DNA experiment, and reads that mapped to chromosomal DNA were removed. Reads sharing identical tag sequences were then grouped together and collapsed to consensus reads. Sequencing positions were discounted if the consensus group covering that position consisted of fewer than three members or if fewer than 90% of the sequences at that position in the consensus group had the identical sequence. A minimum group size of three was selected because next-generation sequencing systems have an average base calling error rate of ~1/100.

Requiring the same base to be identified in three distinct reads decreases the frequency of single-strand consensus sequence (SSCS) errors arising from base-call errors to $(1/100)3 = 1 \times 10^{-6}$, which is below the frequency of spontaneous PCR errors that fundamentally limit the sensitivity of SSCSs. The requirement for 90% of sequences to agree to score a position is a highly conservative cutoff. For example, with a group size of eight, a single disagreeing read will lead to 87.5% agreement and the position will not be scored. If all groups in an experiment are of size nine or less, this cutoff will thus require perfect agreement at any given position to score the position. Further development of our protocol may allow for less stringent parameters to be used to maximize the number of SSCS and duplex consensus sequence (DCS) reads that can be obtained from a given experiment.

Results

[00110] Having established the methodology for Duplex Sequencing with M13mp2 DNA, which is a substrate for which the mutation frequency and spectrum are fairly well established, it was desired to apply the approach to a human DNA sample. Thus, mitochondrial DNA was isolated from human brain tissue and sequenced the DNA after ligation of Duplex Sequencing adapters. A standard sequencing approach with quality filtering for a Phred score of 30 resulted in a mutation frequency of 2.7 × 10−3, and SSCS analysis yielded a mutation frequency of 1.5 × 10−4. In contrast, DCS analysis revealed a much lower overall mutation frequency of 3.5 × 10−5 (Figure 5A). The frequency of mutations in mitochondrial DNA has previously been difficult to measure directly due in part to sources of error in existing assays that can result in either overestimation or underestimation of the true value. An additional confounder has been that most approaches are limited to interrogation of mutations within a small fraction of the genome [56]. The method of single-molecule PCR, which has been proposed as an accurate method of measuring mitochondrial mutation frequency [56] and is considered resistant to damage-induced background errors [57], has resulted in a reported mitochondrial mutation frequency in human colonic mucosa of 5.9 × 10−5 ± 3.2 × 10−5 [56], which is in excellent agreement with our result. Likewise, mitochondrial DNA sequence divergence rates in human pedigrees are consistent with a mitochondrial mutation frequency of 3–5 × 10−5 [58, 59].

[00111] When the distribution of mutations throughout the mitochondrial genome is considered, the quality filtered reads (analyzed without consideration of the tags) have many artifactual errors, such that identification of mutational hotspots is difficult or impossible (Figure 5B). DCS analysis removed these artifacts (Figure 5C) and revealed striking hypermutability of the region of replication initiation (D loop), which is consistent with prior estimates of mutational patterns in mitochondrial DNA based upon sequence variation at this region within the population [60].

[00112] SSCS analysis produced a strong mutational bias, with a 130-fold excess of G→T relative to C→A mutations (Figure 5D), consistent with oxidative damage of the DNA leading to first-round PCR mutations as a significant source of background error. A high level of oxidative damage is expected in mitochondrial DNA, due to extensive exposure of mitochondria to free radical species generated as a byproduct of metabolism [61]. DCS analysis (Figure 5E) removed the mutational bias and revealed that transition mutations are the predominant replication errors in mitochondrial DNA. The DCS mutation spectrum is in accord with prior estimates of deamination events [62] and T-dGTP mispairing by the mitochondrial DNA polymerase [63] as primary mutational forces in mitochondrial DNA. Furthermore, the mutation spectrum of our mitochondrial data are consistent with previous reports of heteroplasmic mutations in human brain showing an increased load of A→G/T→C and G→A/C→T transitions, relative to transversions [64, 65]. A similar spectral bias has also been reported in mice [62, 66] and in population studies of Drosophila melanogaster [67].

Example 3: Demonstration of error-correction by DCS using randomly sheared DNA ends as Single Molecule Identifiers

Materials and Methods

[00113] In addition to those described in the Examples above, the following materials and methods were also used to demonstrate the capability of DCS analysis to remove sequencing errors

[00114] *Sequencing library preparation.* Genomic DNA was isolated from a derivative of *Saccharomyces cerevisiae* strain SC288 by standard methods. The DNA was randomly sheared by the Covaris AFA system, followed by end-repair, A-tailing,

and ligation of Illumina TruSeq DNA sequencing adaptors, all by standard library preparation methods. The resultant sequence data consisted of an average 32.5 fold depth of the 12 megabase *S. cerevisiae* genome.

[00115] *Data analysis.* The first 10 nucleotides of each sequencing read pair, corresponding to the randomly sheared DNA ends, were combined, such that the first 10 nucleotides of read 1, referred to as A, was combined with the first 10 nucleotides of read 2, referred to as B, to yield an SMI tag of form AB. Reads were grouped according to SMI sequence, and nucleotide reads were considered only if they agreed among at least 90% of family members sharing a given tag sequence. For DCS analysis, a tag of form AB1 is partnered with the corresponding tag of form BA2, and nucleotide positions are considered only when the sequence is in agreement among read pairs with matching tags AB1 and BA2.

Results

[00116] In order to demonstrate the capability of DCS analysis to remove sequencing errors, a sequencing library was prepared under standard conditions with commercially available sequencing adaptors, and the randomly sheared DNA ends were used as SMI's. First, reads were grouped by SMI with a minimum family size of 1 member. Considering only sites with a minimum of 20x coverage and with <5% of reads differing from reference, this analysis resulted in 644.8 million nucleotides of sequence data and 2,381,428 mutations, yielding an overall mutation frequency of 3.69×10^{-3}.

[00117] The data was then subjected to DCS analysis with the SMI tags, searching for tags of form AB1 that have partner tags of form BA2, and considered only positions at which the sequence from the two strands was in perfect agreement. 3.1% of the tags had a matching partner present in the library, resulting in 2.9 million nucleotides of sequence data. The sequences of the duplex strands were not complementary at 40,874 nucleotide positions; these disagreeing positions, representing likely sequencing or PCR errors, were removed from analysis. Again considering positions with at least 20x coverage and <5% of reads differing from reference, 3.0 million nucleotides of sequence data and 157 mutations were obtained, with an overall mutation frequency of 5.33×10^{-5}, indicative of removal of >98% of

mutations seen in raw analysis and thereby demonstrating the capability of DCS to lower the error rate of DNA sequencing.

[00118] To compare this result to the method of Kinde et al. [36], reads were grouped into families by SMI tag as before but filtered for families with a minimum of 3 members. This resulted in 1.4 million nucleotides of sequence data and 61 mutations, with an overall mutation frequency of 4.25×10^{-5}. Thus, the method of Kinde et al., with a minimum family size of 3, resulted in less than half as much resultant sequence data after filtering than was obtained by DCS with a minimum family size of 1. Thus, DCS lowered the error rate of sequencing to a comparable degree to a method considered state-of-the-art, but with less loss of sequencing capacity.

Discussion

[00119] It was demonstrated that DCS analysis, using sheared DNA ends as unique molecular identifiers, results in a lowering of the apparent error rate of DNA sequencing. As this proof-of-concept experiment was performed on a library that was not optimized to maximize recovery of both strands, there were not sufficient strand-pairs recovered to perform DCS analysis with a minimum family size of >1 member. Requiring family sizes >1 is expected to further reduce sequencing errors. Moreover, this analysis was limited in that it did not include ligation of degenerate SMI tag sequences; owing to the limited number of shear points flanking any given nucleotide position, use of shear points as SMIs limits the number of unique molecules that can be sequenced in a single experiment. The use of shear points as SMIs in conjunction with an exogenously ligated SMI tag sequence would allow for increased depth of sequencing at any given nucleotide position.

Example 4: Demonstrations of duplex consensus sequencing

[00120] In addition to those described in Examples 1 and 2 above, the following materials and methods were also used.

Materials and Methods

[00121] *Construction of M13mp2 Variants*. M13mp2 gapped DNA encoding the LacZ α fragment was extended by human DNA polymerase δ [2] and the resultant

products were transformed into Escherichia coli and subjected to blue-white color screening as previously described [3]. Mutant plaques were sequenced to determine the location of the mutation resulting in the color phenotype. A series of mutants, each differing from wild type by a single nucleotide change, were then mixed together with wild-type M13mp2 DNA to result in a single final mixture with distinct mutants represented at ratios of 1/10 (G6267A), 1/100 (T6299C), 1/1,000 (G6343A), and 1/10,000 (A6293T).

[00122] *Oxidative Damage of M13mp2 DNA*. Induction of DNA damage was performed by minor modifications to a published protocol [5]: 300 ng of M13mp2 double-stranded DNA was incubated in 10 mM sodium phosphate buffer, pH 7.0, in the presence of 10 μM iron sulfate and 10 μM freshly diluted hydrogen peroxide. Incubation proceeded for 30 min at 37 °C in open 1.5-mL plastic microcentrifuge tubes.

[00123] *DNA Isolation*. M13mp2 DNA was isolated from E. coli strain MC1061 by Qiagen Miniprep. To allow for greater sequencing depth at a defined region of the M13mp2 genome, an 840-bp fragment was enriched by complete digestion with the restriction enzymes Bsu36I and NaeI (New England Biolabs), followed by isolation of the fragment on an agarose gel by the Recochip system (Takara Bio).

Duplex Consensus Sequencing of M13 DNA removes artifactual sequencing errors.

[00124] The spontaneous mutation rate of M13mp2 DNA has been well established by a number of exquisitely sensitive genetic assays to be 3.0E-6 [53], that is, an average of one spontaneous base substitution error for every 330,000 nucleotides. Thus this substrate is well suited as a control for determining the background error frequency of DNA sequencing. M13mp2 DNA was sheared and ligated to adaptors containing double-stranded complementary SMI sequences by standard protocols, and was subjected to deep sequencing on an Illumina HiSeq 2000 followed by Consensus Sequencing analysis (Figure 6).

[00125] Analysis of the data by standard methods (i.e., without consideration of the double stranded SMI sequences) resulted in an error frequency of 3.8E-03, more than one thousand fold higher than the true mutation frequency of M13mp2 DNA. This

indicates that >99.9% of the apparent mutations identified by standard sequencing are in fact artifactual errors.

[00126] The data were then analyzed by Single Strand Consensus Sequencing (SSCS), using the unique SMI tag affixed to each molecule to group PCR products together in order to create a consensus of all PCR products that came from an individual molecule of single-stranded DNA. This resulted in a mutation frequency of 6.4E-OS, suggesting that -98% of sequencing errors are corrected by SSCS.

[00127] Next, the data were subjected to Duplex Consensus Sequencing (DCS), which further corrects errors by using the complementary SMI tags to compare the DNA sequence arising from both of the two strands of a single molecule of duplex DNA. This approach resulted in a mutation frequency of 2.SE-06, in nearly perfect agreement with the true mutation frequency of M13mp2 DNA of 3.0E-06. The number of nucleotides of DNA sequence obtained by a standard sequencing approach, and after SSCS and DCS analysis, may be found in Table 1 below.

[00128] **Table 1**. Data yield from Duplex Sequencing

	M13mp2 DNA	Mitochondrial DNA
Initial nucleotides	6.5×10^9	6.2×10^9
SSCS nucleotides	8.7×10^7	4.1×10^8
DCS nucleotides	$2.2E \times 10^7$	9.7×10^7
Initial reads per SSCS read	75	15
Initial reads per DCS read	295	64
SSCS reads per DCS read	4	4

Initial nucleotides represent raw reads that contain the expected fixed adapter sequence following 12 degenerate nucleotides and map to the reference genome. Apparent nucleotide loss in converting initial reads to SSCSs occurs because many of the initial reads intentionally represent identical PCR duplicates of single-stranded DNA molecules to allow for removal of sequencing and PCR errors by comparison of the sequence among the duplicates. A minimum of three initial reads are required to produce one SSCS; however, a greater average number is necessary to ensure that most DNA fragments have at least this number of duplicates. Under fully optimized conditions, each DCS read would arise from exactly two SSCS reads (one arising from each strand of the initial molecule of duplex DNA). An SSCS:DCS ratio greater than 2 indicates that the strand partner of some SSCSs was not recovered.

[00129] For an artifactual error to be scored by DCS, complementary artifactual errors must occur on both strands of a molecule of duplex DNA. Thus the background (artifactual) error frequency of DCS may be calculated as: (probability of error on one strand) * (probability of error on other strand) * (probability that both errors are complementary).

[00130] As the background error frequency of SSCS in this experiment was -6E-S, the background error frequency of DCS can be calculated as 6E-S * 6E-S * (1/3) = 1.2E-9. This represents a greater than 3 million fold improvement over the error rate of 3.SE-03 that was obtained by a standard sequencing approach.

Consensus sequencing reveals likely sites of DNA damage

[00131] M13mp2 DNA was sequenced as detailed above, with DCS adaptors containing double-stranded complementary SMIs. The spectrum of mutations obtained with SSCS was determined. Data was filtered to consist of forward-mapping reads from Read 1, i.e. sequencing of the reference strand, and reverse-mapping reads from Read 1, i.e. sequencing of the anti-reference strand. True mutations would result in an equal balance between mutations on the reference strand and their complementary mutation on the anti-reference strand.

[00132] However, SSCS analysis revealed a large number of single-stranded G→T mutations on reads mapping in the forward orientation to the reference genome, with a much smaller number of C→A mutations mapping to the reverse orientation. The spectrum of mutations identified by both SSCS and DCS analysis were examined relative to literature reference values [53] for the M13mp2 substrate (Figure 7A). SSCS analysis revealed a large excess of G→A/C→T and G→T/C→A mutations relative to reference ($P < 10^{-6}$, two-sample t test). In contrast, DCS analysis was in excellent agreement with the literature values with the exception of a decrease relative to reference of these same mutational events: G→A/C→T and G→T/C→A ($P < 0.01$). To probe the potential cause of these spectrum deviations, the SSCS data were filtered to consist of forward-mapping reads from read 1 (i.e., direct sequencing of the reference strand) and the reverse complement of reverse-mapping reads from read 1 (i.e., direct sequencing of the antireference strand.) True double-stranded mutations should result in an equal balance of complementary mutations observed on the reference and

antireference strand. However, SSCS analysis revealed a large number of single-stranded G→T mutations, with a much smaller number of C→A mutations (Figure 7B). A similar bias was seen with a large excess of C→T mutations relative to G→A mutations.

[00133] Base-specific mutagenic DNA damage is a likely explanation of these imbalances. Excess G→T mutations are consistent with the oxidative product 8-oxo-guanine (8-oxo-G) causing first round PCR errors and artifactual G→T mutations. DNA polymerases, including those commonly used in PCR, have a strong tendency to insert adenine opposite 8-oxo-G [45, 54], and misinsertion of A opposite 8-oxo-G would result in erroneous scoring of a G→T mutation. Likewise, the excess C→T mutations are consistent with spontaneous deamination of cytosine to uracil [47], a particularly common DNA damage event that results in insertion during PCR of adenine opposite uracil and erroneous scoring of a C→T mutation.

[00134] To determine whether the excess G→T mutations seen in SSCSs might reflect oxidative DNA damage at guanine nucleotides, before sequencing library preparation M13mp2 DNA was incubated with the free radical generator hydrogen peroxide in the presence of iron, a protocol that induces DNA damage [55]. This treatment resulted in a substantial further increase in G→T mutations by SSCS analysis (Figure 8A), consistent with PCR errors at sites of DNA damage as the likely mechanism of this biased mutation spectrum. In contrast, induction of oxidative damage did not alter the mutation spectrum seen with DCS analysis (Figure 8B), indicating that duplex consensus sequences are not similarly susceptible to DNA damage artifacts.

[00135] Furthermore, relative to the literature reference values, DCS analysis results in a lower frequency of G→T/C→A and C→T/G→A mutations (Figure 7A), which are the same mutations elevated in SSCS analysis as a probable result of DNA damage. Notably, the M13mp2 LacZ assay, from which reference values have been derived, is dependent upon bacterial replication of a single molecule of M13mp2 DNA. Thus, the presence of oxidative damage within this substrate could cause an analogous first-round replication error by Escherichia coli, converting a single-stranded damage event into a fixed, double-stranded mutation during replication. The slight reduction in the frequency of these two types of mutations measured by DCS analysis may,

therefore, reflect the absence of damage-induced errors that are scored by the in vivo LacZ assay.

Consensus Sequencing accurately recovers spiked-in control mutations.

[00136] A series of M13mp2 variants were constructed which contain known single base substitutions. These variants were then mixed together at known ratios, and the mixture was prepared for sequencing with DCS adaptors with double-stranded complementary SMIs and was sequenced on an Illumina HiSeq 2000. The data was then analyzed by consensus sequencing (Figure 9). With conventional analysis of the data (i.e. without consideration of the SMI tags), variants present at a level of <1/100 could not be accurately identified. This limitation occurs because at any given position, artifactual mutations are seen at a level of nearly 1/100.

[00137] In contrast, when the data is analyzed by Single Strand Consensus Sequencing (SSCS) with ~20,000 fold depth, accurate recovery of mutant sequence is seen down to one mutant molecule per 10,000 wild type molecules. Duplex Consensus Sequencing (DCS), which was not performed on this sample, would allow for detection of even rarer mutations.

REFERENCES

The references, patents and published patent applications listed below, and all references cited in the specification above are hereby incorporated by reference in their entirety, as if fully set forth herein.

[1] Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31-46.
[2] Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135-45.
[3] Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, et al. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc Natl Acad Sci USA. 2011;108:13177-82.
[4] Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature. 2011;480:368-71.
[5] García-Garcerà M, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, et al. Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. PLoS ONE. 2011;6:e24161.
[6] Fordyce SL, Ávila-Arcos MC, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen FT, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. BioTechniques. 2011;51:127-33.
[7] Druley TE, Vallania FLM, Wegner DJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA. Nat Methods. 2009;6:263-5.
[8] Out AA, van Minderhout IJHM, Goeman JJ, Ariyurek Y, Ossowski S, Schneeberger K, et al. Deep sequencing to reveal new variants in pooled DNA samples. Hum Mutat. 2009;30:1703-12.

[9] Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA. 2008;105:16266-71.
[10] Chiu RWK, Akolekar R, Zheng YWL, Leung TY, Sun H, Chan KCA, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011;342:c7401.
[11] Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008;105:10513-8.
[12] Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2105;481:506-9.
[13] Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. Science Translational Medicine. 2009;1:12ra23-12ra23.
[14] Hyman RW, Herndon CN, Jiang H, Palm C, Fukushima M, Bernstein D, et al. The dynamics of the vaginal microbiome during infertility therapy with in vitro fertilization-embryo transfer. J Assist Reprod Genet. 2012;29:105-15.
[15] LaTuga MS, Ellis JC, Cotton CM, Goldberg RN, Wynn JL, Jackson RB, et al. Beyond bacteria: a study of the enteric microbial consortium in extremely low birth weight infants. PLoS ONE. 2011;6:e27858.
[16] Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: interindividual variation and dynamic response to diet. Genome Res. 2011;21:1616-25.
[17] Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011;49:3463-9.
[18] Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS ONE. 2011;6:e24907.

[19] Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci USA. 2008;105:13081-6.

[20] De Grassi A, Segala C, Iannelli F, Volorio S, Bertario L, Radice P, et al. Ultradeep Sequencing of a Human Ultraconserved Region Reveals Somatic and Constitutional Genomic Instability. PLoS Biol. 2010;8:e1000275.

[21] Zagordi O, Klein R, Däumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Research. 2010;38:7400-9.

[22] Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17:1195-201.

[23] Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, et al. Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat Methods. 2012;9:78-80.

[24] He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464:610-4.

[25] Ameur A, Stewart JB, Freyer C, Hagström E, Ingman M, Larsson N-G, et al. Ultra-Deep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. PLoS Genet. 2011;7:e1002028.

[26] Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng. 2003;96:317-23.

[27] Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. Nucleic Acids Research. 1990;18:1687-91.

[28] Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 2008;5:1005-10.

[29] Salk J, Fox E, Loeb L. Mutational heterogeneity in human cancers: origin and consequences. Annual Review of Pathology. 2009;5:51-75.

[30] Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods. 2009;6:291-5.
[31] Vandenbroucke I, Van Marck H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques. 2011;51:167-77.
[32] Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, et al. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic Acids Research. 2012;40:e2-e.
[33] Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20:273-80.
[34] Miner BE, Stöger RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Research. 2004;32:e135.
[35] McCloskey ML, Stöger R, Hansen RS, Laird CD. Encoding PCR products with batch-stamps and barcodes. Biochem Genet. 2007;45:761-7.
[36] Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci USA. 2011;108:9530-5.
[37] Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA. 2011;108:20166-71.
[38] Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011;9:72-4.
[39] Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. Nucleic Acids Research. 2011;39:e81-e.

[40] Shiroguchi K, Jia TZ, Sims PA, Xie XS. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. Proc Natl Acad Sci USA. 2012;109:1347-52.

[41] Fu GK, Hu J, Wang P-H, Fodor SPA. Counting individual DNA molecules by the stochastic attachment of diverse labels. Proc Natl Acad Sci USA. 2011;108:9026-31.

[42] Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. Embryonic stem cells and somatic cells differ in mutation frequency and type. Proc Natl Acad Sci USA. 2002;99:3586-90.

[43] Lindahl T, Wood RD. Quality control by DNA repair. Science. 1999;286:1897-1905.

[44] Kunkel, TA. Mutational specificity of depurination. Proc Natl Acad Sci USA. 1984;81:1494-98.

[45] Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature. 1991;349:431-4.

[46] Stiller M, Green RE, Ronan M, Simons JF, Du L, He W., et al. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Acad Sci USA. 2006;103:13578-84.

[47] Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obsbet Gynecol. 2011;204:205e1-11.

[48] Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch M, D'Ascenzo M, et al. Whole exome capture in solution with 3 Gbp of data. Genome Biol. 2010;11:R62:1-8.

[49] Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010; 38:159e1-- - 8.

[50] Kaur M, Makrigiorgos GM. Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. Nucleic Acids Res. 2003; 31:26e1-- - 7.

[51] Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature 461, 814–818.

[52] Lynch AM, Sasaki jC, Elespuru R, Jacobson-Kram D, Thybaud V, et al. New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen. 2011; 52(3):205-23.
[53] Thomas DC, Roberts JD, Sabatino RD, Myers TW, et al. Fidelity of mammalian DNA replication and replicative DNA polymerases. Biochemistry. 1991; 30:11751-9.
[54] Kasai H, et al.(1993) Formation, inhibition of formation, and repair of oxidative 8-hydroxyguanine DNA damage. Basic Life Sci 61:257–262.
[55] McBride TJ, Preston BD, Loeb LA (1991) Mutagenic spectrum resulting from DNA damage by oxygen radicals. Biochemistry 30:207–213.
[56] Greaves LC, et al. (2009) Quantification of mitochondrial DNA mutation load. Aging Cell 8:566–572.
[57] Kraytsberg Y, Nicholas A, Caro P, Khrapko K (2008) Single molecule PCR in mtDNA mutational analysis: Genuine mutations vs. damage bypass-derived artifacts. Methods 46:269–273.
[58] Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501–509.
[59] Parsons TJ, et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368.
[60] Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032.
[61] Kennedy SR, Loeb LA, Herr AJ (2011) Somatic mutations in aging, cancer and neurodegeneration. Mech Ageing Dev.
[62] Vermulst M, et al. (2007) Mitochondrial point mutations do not limit the natural lifespan of mice. Nat Genet 39:540–543.
[63] Song S, et al. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. Proc Natl Acad Sci USA 102:4990–4995.
[64] Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Hum Mol Genet 11:133–145.

[65] Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc Natl Acad Sci USA 93:12382–12387.

[66] Khaidakov M, Heflich RH, Manjanatha MG, Myers MB, Aidoo A (2003) Accumulation of point mutations in mitochondrial DNA of aging mice. Mutat Res 526:1–7.

[67] Haag-Liautard C, et al. (2008) Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Biol 6:e204.

ABSTRACT

Next Generation DNA sequencing promises to revolutionize clinical medicine and basic research. However, while this technology has the capacity to generate hundreds of billions of nucleotides of DNA sequence in a single experiment, the error rate of approximately 1% results in hundreds of millions of sequencing mistakes. These scattered errors can be tolerated in some applications but become extremely problematic when "deep sequencing" genetically heterogeneous mixtures, such as tumors or mixed microbial populations. To overcome limitations in sequencing accuracy, a method Duplex Consensus Sequencing (DCS) is provided. This approach greatly reduces errors by independently tagging and sequencing each of the two strands of a DNA duplex. As the two strands are complementary, true mutations are found at the same position in both strands. In contrast, PCR or sequencing errors will result in errors in only one strand. This method uniquely capitalizes on the redundant information stored in double-stranded DNA, thus overcoming technical limitations of prior methods utilizing data from only one of the two strands.

ABSTRACT

Next Generation DNA sequencing promises to revolutionize clinical medicine and basic research. However, while this technology has the capacity to generate hundreds of billions of nucleotides of DNA sequence in a single experiment, the error rate of approximately 1% results in hundreds of millions of sequencing mistakes. These scattered errors can be tolerated in some applications but become extremely problematic when “deep sequencing” genetically heterogeneous mixtures, such as tumors or mixed microbial populations. To overcome limitations in sequencing accuracy, a method Duplex Consensus Sequencing (DCS) is provided. This approach greatly reduces errors by independently tagging and sequencing each of the two strands of a DNA duplex. As the two strands are complementary, true mutations are found at the same position in both strands. In contrast, PCR or sequencing errors will result in errors in only one strand. This method uniquely capitalizes on the redundant information stored in double-stranded DNA, thus overcoming technical limitations of prior methods utilizing data from only one of the two strands.

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

PRIORITY CLAIM

[0001] This application is a continuation of U.S. Patent Application No. 15/660,785, filed July 26, 2017 and now pending, which is a continuation of U.S. Patent Application No. 14/386,800, filed September 20, 2014 and now U.S. Patent No. 9,752,188, which is a U.S. national stage application of International Application No. PCT/US2013/032665, filed March 15, 2013, which claims priority to U.S. Provisional Patent Application No. 61/613,413, filed March 20, 2012; U.S. Provisional Patent Application No. 61/625,623, filed April 17, 2012; and U.S. Provisional Patent Application No. 61/625,319, filed April 17, 2012; the subject matter of all of which are hereby incorporated by reference as if fully set forth herein.

STATEMENT OF GOVERNMENT INTEREST

[0002] This invention was made with government support under Grant Nos. F30AG033485, R01CA102029 and R01CA115802 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] The advent of massively parallel DNA sequencing has ushered in a new era of genomic exploration by making simultaneous genotyping of hundreds of billions of base-pairs possible at small fraction of the time and cost of traditional Sanger methods [1]. Because these technologies digitally tabulate the sequence of many individual DNA fragments, unlike conventional techniques which simply report the average genotype of an aggregate collection of molecules, they offer the unique ability to detect minor variants within heterogeneous mixtures [2].

[0004] This concept of “deep sequencing” has been implemented in a variety fields including metagenomics [3, 4], paleogenomics [5], forensics [6], and human genetics [7, 8] to disentangle subpopulations in complex biological samples. Clinical applications, such prenatal screening for fetal aneuploidy [9, 10], early detection of

cancer [11] and monitoring its response to therapy [12, 13] with nucleic acid-based serum biomarkers, are rapidly being developed. Exceptional diversity within microbial [14, 15] viral [16-18] and tumor cell populations [19, 20] has been characterized through next-generation sequencing, and many low-frequency, drug-resistant variants of therapeutic importance have been so identified [12, 21, 22]. Previously unappreciated intra-organismal mosasism in both the nuclear [23] and mitochondrial [24, 25] genome has been revealed by these technologies, and such somatic heterogeneity, along with that arising within the adaptive immune system [13], may be an important factor in phenotypic variability of disease.

[0005] Deep sequencing, however, has limitations. Although, in theory, DNA subpopulations of any size should be detectable when deep sequencing a sufficient number of molecules, a practical limit of detection is imposed by errors introduced during sample preparation and sequencing. PCR amplification of heterogeneous mixtures can result in population skewing due to stoichastic and non-stoichastic amplification biases and lead to over- or under-representation of particular variants [26]. Polymerase mistakes during pre-amplification generate point mutations resulting from base mis-incorporations and rearrangements due to template switching [26, 27]. Combined with the additional errors that arise during cluster amplification, cycle sequencing and image analysis, approximately 1% of bases are incorrectly identified, depending on the specific platform and sequence context [2, 28]. This background level of artifactual heterogeneity establishes a limit below which the presence of true rare variants is obscured [29].

[0006] A variety of improvements at the level of biochemistry [30-32] and data processing [19, 21, 28, 32, 33] have been developed to improve sequencing accuracy. The ability to resolve subpopulations below 0.1%, however, has remained elusive. Although several groups have attempted to increase sensitivity of sequencing, several limitations remain. For example techniques whereby DNA fragments to be sequenced are each uniquely tagged [34, 35] prior to amplification [36-41] have been reported. Because all amplicons derived from a particular starting molecule will bear its specific tag, any variation in the sequence or copy number of identically tagged sequencing reads can be discounted as technical error. This approach has been used to improve

counting accuracy of DNA [38, 39, 41] and RNA templates [37, 38, 40] and to correct base errors arising during PCR or sequencing [36, 37, 39]. Kinde et. al. reported a reduction in error frequency of approximately 20-fold with a tagging method that is based on labeling single-stranded DNA fragments with a primer containing a 14 bp degenerate sequence. This allowed for an observed mutation frequency of ~0.001% mutations/bp in normal human genomic DNA [36]. Nevertheless, a number of highly sensitive genetic assays have indicated that the true mutation frequency in normal cells is likely to be far lower, with estimates of per-nucleotide mutation frequencies generally ranging from 10^{-9} to 10^{-11} [42]. Thus, the mutations seen in normal human genomic DNA by Kinde et al. are likely the result of significant technical artifacts.

[0007] Traditionally, next-generation sequencing platforms rely upon generation of sequence data from a single strand of DNA. As a consequence, artifactual mutations introduced during the initial rounds of PCR amplification are undetectable as errors - even with tagging techniques - if the base change is propagated to all subsequent PCR duplicates. Several types of DNA damage are highly mutagenic and may lead to this scenario. Spontaneous DNA damage arising from normal metabolic processes results in thousands of damaging events per cell per day [43]. In addition to damage from oxidative cellular processes, further DNA damage is generated *ex vivo* during tissue processing and DNA extraction [44]. These damage events can result in frequent copying errors by DNA polymerases: for example a common DNA lesion arising from oxidative damage, 8-oxo-guanine, has the propensity to incorrectly pair with adenine during complementary strand extension with an overall efficiency greater than that of correct pairing with cytosine, and thus can contribute a large frequency of artifactual G→T mutations [45]. Likewise, deamination of cytosine to form uracil is a particularly common event which leads to the inappropriate insertion of adenine during PCR, thus producing artifactual C→T mutations with a frequency approaching 100% [46].

[0008] It would be desirable to develop an approach for tag-based error correction, which reduces or eliminates artifactual mutations arising from DNA damage, PCR errors, and sequencing errors; allows rare variants in heterogeneous populations to be detected with unprecedented sensitivity; and which capitalizes on the redundant information stored in complexed double-stranded DNA.

SUMMARY

[0009] In one embodiment, a single molecule identifier (SMI) adaptor molecule for use in sequencing a double-stranded target nucleic acid molecule is provided. Said SMI adaptor molecule includes a single molecule identifier (SMI) sequence which comprises a degenerate or semi-degenerate DNA sequence; and an SMI ligation adaptor that allows the SMI adaptor molecule to be ligated to the double-stranded target nucleic acid sequence. The SMI sequence may be single-stranded or double-stranded. In some embodiments, the double-stranded target nucleic acid molecule is a double-stranded DNA or RNA molecule.

[0010] In another embodiment, a method of obtaining the sequence of a double-stranded target nucleic acid is provided (also known as Duplex Consensus Sequencing or DCS) is provided. Such a method may include steps of ligating a double-stranded target nucleic acid molecule to at least one SMI adaptor molecule to form a double-stranded SMI-target nucleic acid complex; amplifying the double-stranded SMI-target nucleic acid complex, resulting in a set of amplified SMI-target nucleic acid products; and sequencing the amplified SMI-target nucleic acid products.

[0011] In some embodiments, the method may additionally include generating an error-corrected double-stranded consensus sequence by (i) grouping the sequenced SMI-target nucleic acid products into families of paired target nucleic acid strands based on a common set of SMI sequences; and (ii) removing paired target nucleic acid strands having one or more nucleotide positions where the paired target nucleic acid strands are non-complementary (or alternatively removing individual nucleotide positions in cases where the sequence at the nucleotide position under consideration disagrees among the two strands). In further embodiments, the method confirms the presence of a true mutation by (i) identifying a mutation present in the paired target nucleic acid strands having one or more nucleotide positions that disagree; (ii) comparing the mutation present in the paired target nucleic acid strands to the error corrected double-stranded consensus sequence; and (iii) confirming the presence of a true mutation when the mutation is present on both of the target nucleic acid strands and appears in all members of a paired target nucleic acid family.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 illustrates an overview of Duplex Consensus Sequencing. Sheared double-stranded DNA that has been end-repaired and T-tailed is combined with A-tailed SMI adaptors and ligated according to one embodiment. Because every adaptor contains a unique, double-stranded, complementary n-mer random tag on each end (n-mer = 12 bp according to one embodiment), every DNA fragment becomes labeled with two distinct SMI sequences (arbitrarily designated α and β in the single capture event shown). After size-selecting for appropriate length fragments, PCR amplification with primers containing Illumina flow-cell-compatible tails is carried out to generate families of PCR duplicates. By virtue of the asymmetric nature of adapted fragments, two types of PCR products are produced from each capture event. Those derived from one strand will have the α SMI sequence adjacent to flow-cell sequence 1 and the β SMI sequence adjacent to flow cell sequence 2. PCR products originating from the complementary strand are labeled reciprocally.

[0013] Figure 2 illustrates Single Molecule Identifier (SMI) adaptor synthesis according to one embodiment. Oligonucleotides are annealed and the complement of the degenerate lower arm sequence (N's) plus adjacent fixed bases is produced by polymerase extension of the upper strand in the presence of all four dNTPs. After reaction cleanup, complete adaptor A-tailing is ensured by extended incubation with polymerase and dATP.

[0014] Figure 3 illustrates error correction through Duplex Consensus Sequencing (DCS) analysis according to one embodiment. (a-c) shows sequence reads (brown) sharing a unique set of SMI tags are grouped into paired families with members having strand identifiers in either the αβ or βα orientation. Each family pair reflects one double-stranded DNA fragment. (a) shows mutations (spots) present in only one or a few family members representing sequencing mistakes or PCR-introduced errors occurring late in amplification. (b) shows mutations occurring in many or all members of one family in a pair representing mutations scored on only one of the two strands, which can be due to PCR errors arising during the first round of amplification such as might occur when copying across sites of mutagenic DNA damage. (c) shows

true mutations (* arrow) present on both strands of a captured fragment appear in all members of a family pair. While artifactual mutations may co-occur in a family pair with a true mutation, these can be independently identified and discounted when producing (d) an error-corrected consensus sequence (i.e., single stranded consensus sequence) (+ arrow) for each duplex. (e) shows consensus sequences from all independently captured, randomly sheared fragments containing a particular genomic site are identified and (f) compared to determine the frequency of genetic variants at this locus within the sampled population.

[0015] Figure 4 illustrates an example of how a SMI sequence with n-mers of 4 nucleotides in length (4-mers) are read by Duplex Consensus Sequencing (DCS) according to some embodiments. (A) shows the 4-mers with the PCR primer binding sites (or flow cell sequences) 1 and 2 indicated at each end. (B) shows the same molecules as in (A) but with the strands separated and the lower strand now written in the 5'-3' direction. When these molecules are amplified with PCR and sequenced, they will yield the following sequence reads: The top strand will give a read 1 file of TAAC--- and a read 2 file of GCCA---. Combining the read 1 and read 2 tags will give TAACCGGA as the SMI for the top strand. The bottom strand will give a read 1 file of CGGA---- and a read 2 file of TAAC---. Combining the read 1 and read 2 tags will give CGGATAAC as the SMI for the bottom strand. (C) illustrates the orientation of paired strand mutations in DCS. In the initial DNA duplex shown in Figures 4A and 4B, a mutation "x" (which is paired to a complementary nucleotide "y") is shown on the left side of the DNA duplex. The "x" will appear in read 1, and the complementary mutation on the opposite strand, "y," will appear in read 2. Specifically, this would appear as "x" in both read 1 and read 2 data, because "y" in read 2 is read out as "x" by the sequencer owing to the nature of the sequencing primers, which generate the complementary sequence during read 2.

[0016] Figure 5 illustrates duplex sequencing of human mitochondrial DNA. (A) Overall mutation frequency as measured by a standard sequencing approach, SSCS, and DCS. (B) Pattern of mutation in human mitochondrial DNA by a standard sequencing approach. The mutation frequency (vertical axis) is plotted for every position

in the ~16-kb mitochondrial genome. Due to the substantial background of technical error, no obvious mutational pattern is discernible by this method. (C) DCS analysis eliminates sequencing artifacts and reveals the true distribution of mitochondrial mutations to include a striking excess adjacent to the mtDNA origin of replication. (D) SSCS analysis yields a large excess of G→T mutations relative to complementary C→A mutations, consistent with artifacts from damaged-induced 8-oxo-G lesions during PCR. All significant ($P < 0.05$) differences between paired reciprocal mutation frequencies are noted. (E) DCS analysis removes the SSCS strand bias and reveals the true mtDNA mutational spectrum to be characterized by an excess of transitions.

[0017] Figure 6 shows that consensus sequencing removes artifactual sequencing errors as compared to Raw Reads. Duplex Consensus Sequencing (DCS) results in an approximately equal number of mutations as the reference and single strand consensus sequencing (SSCS) .

[0018] Figure 7 illustrates duplex sequencing of M13mp2 DNA. (A) Single-strand consensus sequences (SSCSs) reveal a large excess of G→A/C→T and G→T/C→A mutations, whereas duplex consensus sequences (DCSs) yield a balanced spectrum. Mutation frequencies are grouped into reciprocal mispairs, as DCS analysis only scores mutations present in both strands of duplex DNA. All significant ($P < 0.05$) differences between DCS analysis and the literature reference values are noted. (B) Complementary types of mutations should occur at approximately equal frequencies within a DNA fragment population derived from duplex molecules. However, SSCS analysis yields a 15-fold excess of G→T mutations relative to C→A mutations and an 11-fold excess of C→T mutations relative to G→A mutations. All significant ($P < 0.05$) differences between paired reciprocal mutation frequencies are noted.

[0019] Figure 8 shows the effect of DNA damage on the mutation spectrum. DNA damage was induced by incubating purified M13mp2 DNA with hydrogen peroxide and FeSO4. (A) SSCS analysis reveals a further elevation from baseline of G→T mutations, indicating these events to be the artifactual consequence of nucleotide oxidation. All significant ($P < 0.05$) changes from baseline mutation frequencies are noted. (B)

Induced DNA damage had no effect on the overall frequency or spectrum of DCS mutations.

[0020] Figure 9 shows duplex sequencing results in accurate recovery of spiked-control mutations. A series of variants of M13mp2 DNA, each harboring a known single-nucleotide substitution, were mixed in together at known ratios and the mixture was sequenced to ~20,000-fold final depth. Standard sequencing analysis cannot accurately distinguish mutants present at a ratio of less than 1/100, because artifactural mutations occurring at every position obscure the presence of less abundant true mutations, rendering apparent recovery greater than 100%. Duplex consensus sequences, in contrast, accurately identify spiked-in mutations down to the lowest tested ratio of 1/10,000.

[0021] Figure 10 is a Python Code that may used to carry out methods described herein according to one embodiment.

DETAILED DESCRIPTION

[0022] Single molecule identifier adaptors and methods for their use are provided herein. According to the embodiments described herein, a single molecule identifier (SMI) adaptor molecule is provided. Said SMI adaptor molecule is double stranded, and may include a single molecule identifier (SMI) sequence, and an SMI ligation adaptor (Figure 2). Optionally, the SMI adaptor molecule further includes at least two PCR primer binding sites, at least two sequencing primer binding sites, or both.

[0023] The SMI adaptor molecule may form a "Y-shape" or a "hairpin shape." In some embodiments, the SMI adaptor molecule is a "Y-shaped" adaptor, which allows both strands to be independently amplified by a PCR method prior to sequencing because both the top and bottom strands have binding sites for PCR primers FC1 and FC2 as shown in the examples below. A schematic of a Y-shaped SMI adaptor molecule is also shown in Figure 2. A Y-shaped SMI adaptor requires successful amplification and recovery of both strands of the SMI adaptor molecule. In one embodiment, a modification that would simplify consistent recovery of both strands entails ligation of a Y-shaped SMI adaptor molecule to one end of a DNA duplex molecule, and ligation of a "U-shaped" linker to the other end of the molecule. PCR

amplification of the hairpin-shaped product will then yield a linear fragment with flow cell sequences on either end. Distinct PCR primer binding sites (or flow cell sequences FC1 and FC2) will flank the DNA sequence corresponding to each of the two SMI adaptor molecule strands, and a given sequence seen in Read 1 will then have the sequence corresponding to the complementary DNA duplex strand seen in Read 2. Mutations are scored only if they are seen on both ends of the molecule (corresponding to each strand of the original double-stranded fragment), i.e. at the same position in both Read 1 and Read 2. This design may be accomplished as described in the examples relating to double stranded SMI sequence tags.

[0024] In other embodiments, the SMI adaptor molecule is a "hairpin" shaped (or "U-shaped") adaptor. A hairpin DNA product can be used for error correction, as this product contains both of the two DNA strands. Such an approach allows for reduction of a given sequencing error rate N to a lower rate of N*N*(1/3), as independent sequencing errors would need to occur on both strands, and the same error among all three possible base substitutions would need to occur on both strands. For example, the error rate of 1/100 in the case of Illumina sequencing [32] would be reduced to (1/100)*(1/100)*(1/3) = 1/30,000.

[0025] An additional, more remarkable reduction in errors can be obtained by inclusion of a single-stranded SMI in either the hairpin adaptor or the "Y-shaped" adaptor will also function to label both of the two DNA strands. Amplification of hairpin-shaped DNA may be difficult as the polymerase must synthesize through a product containing significant regions of self-complementarity, however, amplification of hairpin-shaped structures has already been established in the technique of hairpin PCR, as described below. Amplification using hairpin PCR is further described in detail in U.S. Patent No. 7,452,699, the subject matter of which is hereby incorporated by reference as if fully set forth herein.

[0026] According to the embodiments described herein, the SMI sequence (or "tag") may be a double-stranded, complementary SMI sequence or a single-stranded SMI sequence. In some embodiments, the SMI adaptor molecule includes an SMI sequence (or "tag") of nucleotides that is degenerate or semi-degenerate. In some

embodiments, the degenerate or semi-degenerate SMI sequence may be a random degenerate sequence. A double-stranded SMI sequence includes a first degenerate or semi-degenerate nucleotide n-mer sequence and a second n-mer sequence that is complementary to the first degenerate or semi-degenerate nucleotide n-mer sequence, while a single-stranded SMI sequence includes a first degenerate or semi-degenerate nucleotide n-mer sequence. The first and/or second degenerate or semi-degenerate nucleotide n-mer sequences may be any suitable length to produce a sufficiently large number of unique tags to label a set of sheared DNA fragments from a segment of DNA. Each n-mer sequence may be between approximately 3 to 20 nucleotides in length. Therefore, each n-mer sequence may be approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 nucleotides in length. In one embodiment, the SMI sequence is a random degenerate nucleotide n-mer sequence which is 12 nucleotides in length. A 12 nucleotide SMI n-mer sequence that is ligated to each end of a target nucleic acid molecule, as described in the Example below, results in generation of up to 4^{24} (i.e., 2.8×10^{14}) distinct tag sequences.

[0027] In some embodiments, the SMI tag nucleotide sequence may be completely random and degenerate, wherein each sequence position may be any nucleotide. (i.e., each position, represented by "X," is not limited, and may be an adenine (A), cytosine (C), guanine (G), thymine (T), or uracil (U)) or any other natural or non-natural DNA or RNA nucleotide or nucleotide-like substance or analog with base-pairing properties (e.g., xanthosine, inosine, hypoxanthine, xanthine, 7-methylguanine, 7-methylguanosine, 5,6-dihydrouracil, 5-methylcytosine, dihydouridine, isocytosine, isoguanine, deoxynucleosides, nucleosides, peptide nucleic acids, locked nucleic acids, glycol nucleic acids and threose nucleic acids). The term "nucleotide" as described herein, refers to any and all nucleotide or any suitable natural or non-natural DNA or RNA nucleotide or nucleotide-like substance or analog with base pairing properties as described above. In other embodiments, the sequences need not contain all possible bases at each position. The degenerate or semi-degenerate n-mer sequences may be generated by a polymerase-mediated method described in the Example below, or may be generated by preparing and annealing a library of individual oligonucleotides of known sequence. Alternatively, any degenerate or semi-degenerate n-mer sequences

may be a randomly or non-randomly fragmented double stranded DNA molecule from any alternative source that differs from the target DNA source. In some embodiments, the alternative source is a genome or plasmid derived from bacteria, an organism other than that of the target DNA, or a combination of such alternative organisms or sources. The random or non-random fragmented DNA may be introduced into SMI adaptors to serve as variable tags. This may be accomplished through enzymatic ligation or any other method known in the art.

[0028] In some embodiments, the SMI adaptor molecules are ligated to both ends of a target nucleic acid molecule, and then this complex is used according to the methods described below. In certain embodiments, it is not necessary to include n-mers on both adapter ends, however, it is more convenient because it means that one does not have to use two different types of adaptors and then select for ligated fragments that have one of each type rather than two of one type. The ability to determine which strand is which is still possible in the situation wherein only one of the two adaptors has a double-stranded SMI sequence.

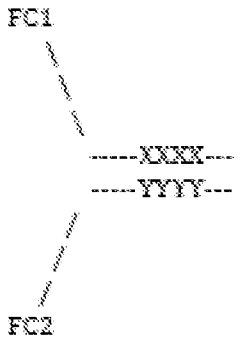
[0029] In some embodiments, the SMI adaptor molecule may optionally include a double-stranded fixed reference sequence downstream of the n-mer sequences to help make ligation more uniform and help computationally filter out errors due to ligation problems with improperly synthesized adaptors. Each strand of the double-stranded fixed reference sequence may be 4 or 5 nucleotides in length sequence, however, the fixed reference sequence may be any suitable length including, but not limited to 3, 4, 5 or 6 nucleotides in length.

[0030] The SMI ligation adaptor may be any suitable ligation adaptor that is complementary to a ligation adaptor added to a double-stranded target nucleic acid sequence including, but not limited to a T-overhang, an A-overhang, a CG overhang, a blunt end, or any other ligatable sequence. In some embodiments, the SMI ligation adaptor may be made using a method for A-tailing or T-tailing with polymerase extension; creating an overhang with a different enzyme; using a restriction enzyme to create a single or multiple nucleotide overhang, or any other method known in the art.

[0031] According to the embodiments described herein, the SMI adaptor molecule may include at least two PCR primer or "flow cell" binding sites: a forward PCR primer binding site (or a "flow cell 1" (FC1) binding site); and a reverse PCR primer binding site (or a "flow cell 2" (FC2) binding site). The SMI adaptor molecule may also include at least two sequencing primer binding sites, each corresponding to a sequencing read. Alternatively, the sequencing primer binding sites may be added in a separate step by inclusion of the necessary sequences as tails to the PCR primers, or by ligation of the needed sequences. Therefore, if a double-stranded target nucleic acid molecule has an SMI adaptor molecule ligated to each end, each sequenced strand will have two reads - a forward and a reverse read.

Double-stranded SMI sequences

[0032] Adaptor 1 (shown below) is a Y-shaped SMI adaptor as described above (the SMI sequence is shown as X's in the top strand (a 4-mer), with the complementary bottom strand sequence shown as Y's):

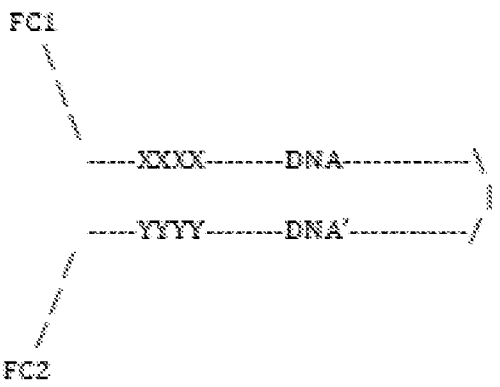


(Adaptor 1)

[0033] Adaptor 2 (shown below) is a hairpin (or "U-shaped") linker:

(Adaptor 2)

[0034] Following ligation of both adaptors to a double-stranded target nucleic acid, the following is structure is obtained:



[0035] When melted, the product will be of the following form (where "linker" is the sequence of adaptor 2):

FC1-------XXXX-----DNA---------linker---------DNA'----------YYYY--------FC2

[0036] This product is then PCR amplified. The reads will yield:

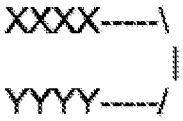
Read 1:
XXXX-----DNA----

Read 2 (note that read 2 is seen as the complement of the bases sequenced:)
XXXX-----DNA----

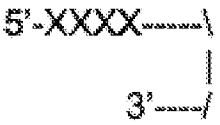
[0037] The sequences of the two duplex strands seen in the two sequence reads may then be compared, and sequence information and mutations will be scored only if the sequence at a given position matches in both of the reads.

[0038] This approach does not strictly require the use of an SMI tag, as the sheared ends can be used as identifiers to differentiate unique individual molecules from PCR duplicates. Thus the same concept would apply if one used any standard sequencing adaptor as "Adaptor 1" and the U---shaped linker as "Adaptor 2." However described below, there are a limited number of shear points flanking any given genomic position and thus the power to sequence deeply is increased via inclusion of the SMI tag. A hybrid method using a combination of sheared ends and a shorter n-mer tag (such as 1 or 2 or 3 or 4 or more degenerate or semi-degenerate bases) in the adaptor may also serve as unique molecular identifiers. Another design may include use of any sequencing adaptor (such as one lacking an n-mer tag) in conjunction with an n-mer tag that is incorporated into the U-shaped linker molecule. Such a design would be of the

following form (where X and Y represent complementary degenerate or semi-degenerate nucleotides):

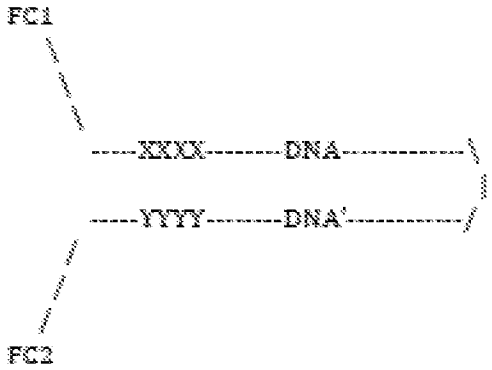


[0039] Synthesis of such a design may be obtained in a number of ways, for example synthesizing a set of hairpin oligonucleotides in which each individual oligonucleotide encodes a complementary n-mer sequence, or alternatively by using a DNA polymerase to carry out extension from the following product (where X's represent degenerate nucleotides):



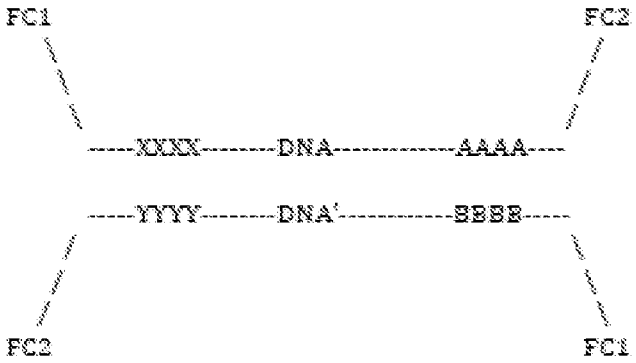
[0040] Inclusion of the SMI tag is also extremely useful for identifying correct ligation products, as the assay uses two distinct adaptors. This will yield multiple possible ligation products:

[0041] **Product I.** Adaptor 1---------DNA---------Adaptor 2, which yields the desired product:

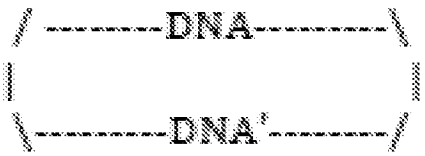


[0042] **Product II.** Adaptor 1---------DNA---------Adaptor 1. This will result in the DNA being amplified as two separate strands, i.e. as occurs in the DCS approach described elsewhere in this document (the second copy of Adaptor 1 is shown below

with the SMI as AAA – BBB to emphasize that every DCS adaptor has a distinct SMI sequence)



[0043] **Product III.** Adaptor 2---------DNA---------Adaptor 2. This will result in a non---amplifiable circular product shown below:



[0044] Product III is non-amplifiable, given the absence of primer binding sites and thus will not be present in the final DNA sequences. Thus only Product II needs to be avoided. The formation of Product II can be minimized in the ligation step by using an excess of Adaptor 2 (relative to Adaptor 1). Then primarily Products I and III will be obtained, with minimal formation of Product II. Additionally, a variety of biochemical means of enriching for products containing adaptor 2 are possible such as using affinity probes that are complementary to the hairpin loop sequence itself. Product I results in the same SMI sequence in both the Read 1 and Read 2 sequence reads. In the example depicted above, Product I sequences can thus be identified by virtue of having matching SMIs of the form XXXX in Read 1 and XXXX in Read 2.

[0045] By contrast, in the case of Product II, the SMI sequences on either end of the sequenced molecule will arise from distinct DCS adaptors having different SMI sequences. In the example shown above, Product II sequences yield SMIs of the form XXXX (Read 1) – BBB (Read 2) upon sequencing of the top strand, and BBBB (Read 1)

– XXXX (Read 2) upon sequencing of the bottom strand. Thus Product II sequences can be easily identified and computationally removed from the final sequence data.

[0046] Data resulting from Product II is useful, because Product II corresponds to the product analyzed under the approach detailed in the Example below. Product I contains a self--complementary hairpin sequence that can impair polymerase extension during amplification, however, this type of amplification has already been enabled in the technique of "Hairpin PCR" [50] which involves linking of the two strands followed by amplification with gene-specific primers. Amplification conditions that are compatible with amplification of hairpin DNA are thus already established. Moreover, ligation and amplification with circularizing "linkers" (i.e. hairpin linkers affixed to both ends of a fragment) has been demonstrated as a step in the Pacific Biosciences sample preparation workflow [49]. As the sequence of the linker itself does not matter in the workflow, the published linker sequences from either of these references would be adequate for use in the assay.

[0047] In some aspects of some embodiments, deliberate ligation of "U-shaped" adaptors or hairpin linkers containing 1) a double-stranded n-mer (or other form of degenerate or semi-degenerate double-stranded tag as enumerated above) plus 2) primer binding sites to both ends of a captured fragment may be desireable. Producing closed circles of captured material may help facilitate removal of non-captured DNA by exonuclease digestion given that circularized DNA will be protected from digestion by such enzymes. Additionally, closed circles may be pre-amplified using rolling circle amplification or serve as the substrate for continuous loop sequencing [49]. Recognition sites for restriction endonuclease digestion could be engineered into these adaptors to render closed loops open once again if more convenient for subsequent steps.

[0048] In another embodiment, flow cell sequences or PCR binding sites, again denoted as FC1 and FC2, may be included in both the PCR primers and the hairpin linker adaptor, as well as a ligatable sequence on the end of the hairpin linker (denoted as L below). The hairpin linker adaptor may additionally include one or more cleavable sequences, denoted as R in the example below (the R may be any appropriate

restriction enzyme target sequence, or any other cleavable sequence). Such a hairpin linker design is shown below:

```
L-XXXX--FC2-R----\
                  |
L-YYYY--FC1-R----/
```

[0049] The target DNA with ligation site denoted as L is as follows:

```
--------------DNA--------L
--------------DNA'-------L
```

[0050] Following ligation of the linker, the product may be amplified with PCR primers as follows:

```
FC1
   \
    \
     \
      ------
      --------------DNA-------- XXXX--FC2-R----\
                                                |
      --------------DNA'------- YYYY--FC1-R----/
            ------
          /
         /
        /
      FC2
```

[0051] The resultant product will be of the form:

```
FC1--------DNA--------XXXX--FC2-R-----R-FC1--YYYY--------DNA'------- FC2
```

[0052] After amplification of the product, the cleavage sites R may be cleaved to result in the following sequencable products:

```
FC1--------DNA-------- XXXX--FC2
```

and

```
FC1--YYYY--------DNA'------- FC2
```

[0053] These products may then be sequenced directly. This design has the advantage of allowing for targeted sequencing of a specific region of the genome, and

furthermore avoids the need to sequence a hairpin product, as sequencing of a hairpin will be less efficient due to the self-complementarity present within the hairpin molecule.

Single-stranded SMI sequences

[0054] In one embodiment, a single-stranded SMI sequence is incorporated into the single-stranded portion of the hairpin loop (regions of sequence complementarity are denoted as "="). The SMI sequence is shown as four nucleotides in length in the following examples, but in practice an Nmer of any length, including approximately 3 to 20 nucleotides, will suffice.

```
=====NNNN-----\
              |
=====---------/
```

[0055] Ligation of the hairpin linker and a Y-shaped sequencing adaptor (with PCR primer binding sites labeled as FC1 and FC2) yields the following product:

```
 FC1
    \
     \
      =====DNA =====NNNN------\
                             |
      =====DNA'=====----------/
     /
    /
   /
FC2
```

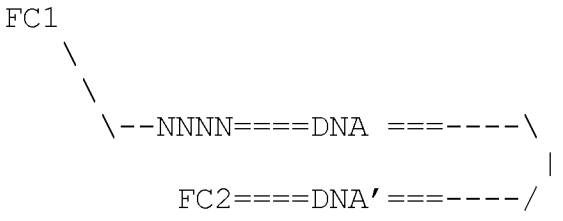
[0056] Melting and PCR amplification of this product yields the following DNA product:

```
FC1-------DNA-----NNNN------hairpin sequence------DNA'-------FC2
```

[0057] Following PCR duplication of the product and formation of consensus reads based upon the shared SMI sequence among all the PCR duplicates, the sequences of the two strands (denoted DNA and DNA') can then be compared to form a duplex consensus sequence.

[0058] In another embodiment, a single-stranded SMI is incorporated into a modified "Y-shaped" sequencing adaptor in which PCR primer binding sites are located

at the sites labeled FC1 and FC2 (regions of sequence complementarity are depicted as "=")



[0059] It will be apparent to one skilled in the art that a single-stranded SMI sequence tag can be located in any of several positions within either the sequencing adaptor or the hairpin linker. The single-stranded SMI sequence tag can be synthesized as a random oligonucleotide sequence, or can be sequenced as a set of fixed sequences by synthesis on an array, or by any other suitable method known in the art.

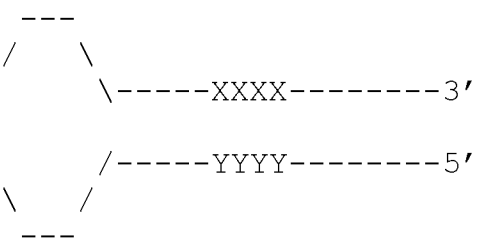
Methods for synthesis of complementary or partially complementary double stranded SMI tags

[0060] SMI adaptors molecules containing a double-stranded, complementary, degenerate or semi-degenerate SMI tag can be made by any of a number of methods, including copying of a single-stranded SMI sequence by a DNA polymerase as described above or synthesis and annealing of two oligonucleotides containing complementary SMI sequences. An additional method involves synthesizing a set of linear oligonucleotides which will self-anneal into the appropriate form. Inclusion of a cleavable linker in each oligonucleotide will then allow for conversion of a "hairpin shaped" SMI adaptor molecule into a "Y-shaped" SMI adaptor molecule. For example, an oligonucleotide may be prepared of the following form:

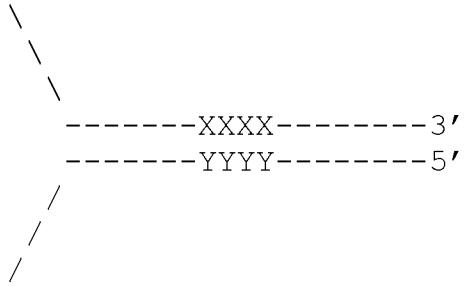
5'---------YYYY-------U-------XXXX----------3'

[0061] In this schematic, X and Y represent complementary nucleotides, and U indicates a cleavable linker, such as uracil (which can be cleaved by combined treatment with uracil DNA glycosylase and apurinic endonuclease), although any other

cleavable linker will suffice. The oligonucleotide may be designed with appropriate regions of self-complementarity to anneal into the following form:



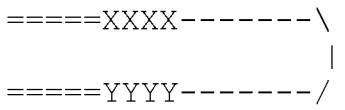
[0062] The linker (e.g. uracil) may then be cleaved, yielding a DCS adaptor:



[0063] A double-stranded SMI hairpin linker can be constructed by an analogous method but without the need for a cleavable linker. For example, a set of nucleotides of known sequence where X and Y represent the complementary SMI sequences can be synthesized on an array, or by any other suitable method known in the art:

5′=====XXXX-------------YYYY=====3′

[0064] This oligonucleotide can then self-anneal to form a hairpin linker with complementary SMI sequences.



[0065] Any of the oligonucleotides described above can also include any ligatable sequence as overhangs on either the 5' or 3' end, or can be used for blunt end ligation.

DCS SMI adaptor molecules may include sequences to allow for targeted DNA capture

[0066] DCS SMI adaptor molecules contain ligatable ends to allow attachment of the adaptor to a target DNA molecule. In some embodiments, the ligatable end may be complementary to a DNA overhang on the target DNA, for example, one generated by digestion of target DNA with a restriction endonuclease. Selective ligation of the adaptor to the targeted DNA containing the matching Single-stranded overhanging DNA sequence will then allow for partial purification of the targeted DNA. A non-limiting example of this embodiment is shown above. In some embodiments, the DCS SMI adaptor molecule, or a hairpin linker SMI adaptor molecule, may additionally contain modifications such as biotin to facilitate affinity purification of target DNA that has ligated to the adaptor.

[0067] In another embodiments, specific PCR primers can selectively amplify specific regions of genome when the adaptor that is ligated to the other end of the molecule is a hairpin (or “U-shape”). Alternatively, this method may be used with or without the need for this cleavable hairpin sequence.

Preparation of DNA for duplex consensus sequencing may be performed by PCR amplification in a hairpin structure

[0068] Another embodiment involves fragmentation of DNA at defined regions, for example by treatment of DNA with a site-specific restriction endonuclease or a mixture of such endonucleases, followed by annealing of a hairpin oligonucleotide linker, and amplification of the hairpin complex with PCR primers sufficient for amplification of the desired DNA sequence. Annealing of the hairpin linker to only one of the two ends of the DNA duplex could be accomplished by using different restriction enzymes to cut on either end of the target duplex, and then having the hairpin linker ligation adaptor being ligatable to only one of the two resultant ligatable ends.

[0069] The example shown below indicates forward and reverse PCR primers (labeled 1 and 2) in conjunction with a hairpin linker to allow linked amplification of both complementary strands of duplex DNA. Such amplification, in conjunction with a single-stranded or double-stranded SMI sequence, would allow for targeted amplification and high accuracy deep sequencing of a specific sequence of interest. In the schematic shown below, a single-stranded SMI sequence is incorporated into PCR primer FC1. It would be apparent to one skilled in the art that the SMI sequence could also be incorporated in primer FC2, or in the hairpin linker.

```
FC1NNNN___
         DNA ==========\
                       |
         DNA'==========/
      FC2___
```

[0070] Amplified product:

```
FC1NNNN DNA----hairpin sequence----DNA'FC2
```

[0071] This product can then be subjected to consensus sequencing analysis. The SMI sequence allows one to group together products of PCR amplification arising from a single molecule of duplex DNA. The sequences of the two DNA strands can then be compared for error correction.

Uses of SMI adapter molecules

[0072] The SMI adaptor molecules described herein have several uses. In some embodiments, the SMI adaptor molecules described herein may be used in methods to obtain the sequence or other sequence-related information of a double-stranded target nucleic acid molecule. According to the embodiments described herein, the term "double-stranded target nucleic acid molecule" includes a double-stranded DNA molecule or a double-stranded RNA molecule. Thus, the SMI adaptor molecules and methods of use described herein are applicable to genotyping and other applications related to sequencing of DNA molecules, but are also applicable to RNA sequencing applications such as for sequencing of double-stranded RNA viruses. Methods for sequencing RNA may include any of the embodiments described herein with respect to

DNA sequencing, and vice-versa. For example, any double stranded target nucleic acid molecule may be ligated to an SMI adaptor molecule which includes a double-stranded RNA or DNA n-mer tag and an RNA or DNA ligation adapter as described above. Methods exist for directly sequencing RNA [51]; alternatively, the ligated product may be reverse transcribed into DNA, and then sequenced as a double-stranded target DNA molecule.

[0073] In one embodiment, the double-stranded target nucleic acid molecule may be a sheared double-stranded DNA or RNA fragment. The sheared target DNA or RNA molecule may be end repaired and a double-stranded target nucleic acid sequence ligation adaptor may be added to each end of the sheared target DNA or RNA molecule. The double-stranded target nucleic acid sequence ligation adaptor may be any suitable ligation adaptor that is complementary to the SMI ligation adaptor described above including, but not limited to a T-overhang, an A-overhang, a CG overhang, blunt end or any other ligatable sequence. In some embodiments, the double-stranded target nucleic acid sequence ligation adaptor may be made using a method for A-tailing or T-tailing with polymerase extension; adding an overhang with a different enzyme; using a restriction enzyme to create a ligatable overhang; or any other method known in the art.

[0074] Methods to obtain the sequence or other sequence-related information of a double-stranded target nucleic acid molecule may include a step of ligating the double-stranded target nucleic acid molecule to at least one SMI adaptor molecule, such as those described above, to form a double-stranded target nucleic acid complex. In one embodiment, each end of the double-stranded target nucleic acid molecule is ligated to an SMI adaptor molecule. The double-stranded target nucleic acid complex is then amplified by a method known in the art (e.g., a PCR or non-PCR method known in the art), resulting in a set of uniquely labeled, amplified SMI-target nucleic acid products. These products are then sequenced using any suitable method known in the art including, but not limited to, the Illumina sequencing platform, ABI SOliD sequencing platform, Pacific Biosciences sequencing platform, 454 Life Sciences sequencing platform, Ion Torrent sequencing platform, Helicos sequencing platform, and nanopore sequencing technology.

[0075] In certain embodiments, a method of generating an error corrected double-stranded consensus sequence is provided. Such a method, also referred to as duplex consensus sequencing (DCS), allows for a quantitative detection of sites of DNA damage. DCS analysis facilitates the detection of DNA damage signatures, in that single stranded DNA mutations that are not present in the complementary strand can be inferred to be artifactual mutations arising from damaged nucleotides. Not only can one correct for these erroneous mutations, but the ability to indirectly infer that damage is present on the DNA could be a useful biomarker (e.g. for cancer risk, cancer metabolic state, mutator phenotype related to defective damage repair, carcinogen exposure, chronic inflammation exposure, individual-specific aging, neurodegenerative diseases etc). The ability to use different polymerases during the first round(s) of PCR to mis-incorporate at damage sites could potentially add even more information. Besides polymerases, other DNA modifying/repair enzymes could be used prior to amplification to convert damage of one sort that doesn't give a specific mutagenic signature into another sort that does with whatever polymerase is used. Alternatively, DNA modifying/repair enzymes could be used to remove damaged bases, and one could sequence both strands of DNA both with and without the enzymatic treatment. Mutations in single-stranded DNA that are seen to be removed by the enzymatic treatment can thus be inferred to be arising due to DNA damage. This could be useful on human nuclear or mtDNA but also might also be useful with model organisms (mice, yeast, bacteria etc), treated with different new damaging agents, facilitating a screen for DNA damaging compounds that would be analogous to the widely used Ames test [52].

[0076] The method of generating an error corrected double-stranded consensus sequence may include a first stage termed “single strand consensus sequencing” (SSCS) followed by a second stage of duplex consensus sequencing (DCS). Therefore, the method includes steps of tagging individual duplex DNA molecules with an SMI adaptor molecule, such as those described above; generating a set of PCR duplicates of the tagged DNA molecules by performing a suitable PCR method; creating a single strand consensus sequence from all of the PCR duplicates which arose from an individual molecule of single-stranded DNA. Each DNA duplex should result in two

single strand consensus sequences. The work through these three steps conclude the first stage and is termed SSCS.

[0077] The method of generating an error corrected double-stranded consensus sequence further comprises the second stance that is termed DCS. The DCS stage includes steps of comparing the sequence of the two single strand consensus sequences arising from a single duplex DNA molecule, and further reducing sequencing or PCR errors by considering only sites at which the sequences of both single-stranded DNA molecules are in agreement. The method that includes the first stage and the second stage termed Duplex Consensus Sequencing (DCS).

[0078] The step of tagging of both strands of individual duplex DNA may be accomplished by ligation of degenerate or semi-degenerate complementary DNA sequences; as the complementary nature of the two strands of such a tag sequence allows the two molecules to be grouped together for error correction. Alternatively, as described above, the two duplex DNA strands may be linked by ligation of a U-shaped SMI adaptor molecule, and the two DNA strands can thus both be tagged with a single-stranded SMI tag.

[0079] In the method described above, a set of sequenced SMI-DNA products generated in the methods described above may be grouped into families of paired target nucleic acid strands based on a common set of SMI sequences. Then, the paired target nucleic acid strands can be filtered to remove nucleotide positions where the sequences seen on both of the paired partner DNA strands are not complementary. This error corrected double-stranded consensus sequence may be used in a method for confirming the presence of a true mutation (as opposed to a PCR error or other artifactual mutation) in a target nucleic acid sequence. According to certain embodiments, such a method may include identifying one or more mutations present in the paired target nucleic acid strands that have one or more nucleotide positions that disagree between the two strands, then comparing the mutation present in the paired target nucleic acid strands to the error corrected double-stranded consensus sequence. The presence of a true mutation is confirmed when the mutation is present on both of

the target nucleic acid strands and also appear in all members of a pared target nucleic acid family.

[0080] The accuracy of current approaches to next-generation sequencing is limited due to their dependence on interrogating single-stranded DNA. This dependence makes potential sources of error such as PCR amplification errors and DNA damage fundamentally limiting. However, the complementary strands of a double-stranded DNA molecule (or "DNA duplex") contain redundant sequencing information (i.e., one molecule reciprocally encoding the sequence information of its partner) which can be utilized to eliminate such artifacts. Limitations related to sequencing single-stranded DNA (e.g., sequencing errors) may therefore be overcome using the methods described herein. This is accomplished by individually tagging and sequencing each of the two strands of a double-stranded (or duplex) target nucleic acid molecule and comparing the individual tagged amplicons derived from one half of a double-stranded complex with those of the other half of the same molecule. Duplex Consensus Sequencing (DCS), significantly lowers the error rate of sequencing. In some embodiments, the DCS method may be used in methods for high sensitivity detection of rare mutant and variant DNA as described further below.

[0081] As described above, one approach that has previously been reported for DNA sequencing involves incorporation of a random tag sequence into a PCR primer [36]. This approach results in an improvement in accuracy relative to standard Illumina sequencing, but is fundamentally limited in that it is based upon amplification and sequencing of single-stranded DNA and thus cannot overcome limitations in sensitivity owing to single-stranded DNA damage events. In the methods described herein, PCR duplicates are generated from a single strand of DNA, and the sequences of the duplicates are compared. Mutations are scored only when they are present in multiple replicates of a single starting molecule. The DCS approach overcomes the limitation of previous approaches by considering both DNA strands.

[0082] DNA damage should not be a limiting factor in DCS, because miscoding damage events at a single base-pair position occur essentially exclusively on only one of the two DNA strands. For DNA damage to result in an artifactual mutation in DCS,

damage would need to be present at the same nucleotide position on both strands. Even if complementary nucleotides in a duplex were both damaged, the damage would need to result in complementary sequencing errors to result in mis-scoring of a mutation. Likewise, spontaneous PCR errors would need to result in complementary mutations at the same position on both strands; with a first-round mutation frequency of Taq polymerase of approximately 10^{-5} and three possible incorrect bases that could be mis-inserted, the probability of two complementary PCR errors occurring would be $10^{-5} \times 10^{-5} \times 1/3 = 3.3 \times 10^{-11}$

[0083] According to some embodiments, the sequencing method may be performed using the Illumina or similar platforms including those enumerated above without the use of SMI adaptor molecules, but instead by using the random shear points of DNA as identifiers. For a given DNA sequence seen in sequencing read 1 with a specific set of shear points, the partner strand will be seen as a matching sequence in read two with identical shear points. In practice, this approach is limited by the limited number of possible shear points that overlap any given DNA position. However, according to some embodiments, shear points of a target nucleic acid molecule may be used as unique identifiers to identify double-stranded (or duplex) pairs, resulting in an apparent error frequency at least as low as that seen with traditional sequencing methods, but with a significantly lower loss of sequence capacity. In other embodiments, DCS based on shear points alone may have a role for confirmation that specific mutations of interest are true mutations which were indeed present in the starting sample (i.e. present in both DNA strands), as opposed to being PCR or sequencing artifacts. Overall, however, DCS is most generally applicable when randomized, complementary double-stranded SMI sequences are used. A 24 nucleotide double-stranded SMI sequence was used in the Example described below, which may yield up to $4^{24} = 2.8 \times 10^{14}$ distinct double-stranded SMI sequences. Combining information regarding the shear points of DNA with the SMI tag sequence would allow a shorter SMI to be used, thus minimizing loss of sequencing capacity due to sequencing of the SMI itself.

[0084] In certain embodiments, the SMI adaptor molecules may also be used in methods of single-molecule counting for accurate determination of DNA or RNA copy

number [38]. Again, since the SMI tags are present in the adaptors, there are no altered steps required in library preparation, which is in contrast to other methods for using random tags for single-molecule counting. Single-molecule counting has a large number of applications including, but not limited to, accurate detection of altered genomic copy number (e.g., for sensitive diagnosis of genetic conditions such as trisomy 21 [47]), for accurate identification of altered mRNA copy number in transcriptional sequencing and chromatin immunoprecipitation experiments, quantification of circulating microRNAs, quantification of viral load of DNA or RNA viruses, quantification of microorganism abundance, quantification of circulating neoplastic cells, counting of DNA-labeled molecules of any variety including tagged antibodies or aptamers, and quantification of relative abundances of different individual's genomes in forensic applications.

[0085] In another embodiment, the SMI adaptor molecules may be used in methods for unambiguous identification of PCR duplicates. In order to restrict sequencing analysis to uniquely sequenced DNA fragments, many sequencing studies include a step to filter out PCR duplicates by using the shear points at the ends of DNA molecules to identify distinct molecules. When multiple molecules exhibit identical shear points, all but one of the molecules are discarded from analysis under the assumption that the molecules represent multiple PCR copies of the same starting molecule. However sequence reads with identical shear points can also reflect distinct molecules because there are a limited number of possible shear points at any given genomic location, and with increasing sequencing depth, recurrent shear points are increasingly likely to be seen [48]. Because the use of SMI tags (or "double-stranded SMI sequences") allows every molecule to be uniquely labeled prior to PCR duplication, true PCR duplicates may be unambiguously identified by virtue of having a common (i.e., the same or identical) SMI sequence. This approach would thereby minimize the loss of data by overcoming the intrinsic limitations of using shear points to identify PCR duplicates.

[0086] Importantly, once SMI-containing adaptors are synthesized by a straightforward series of enzymatic steps or are produced through synthesis of a set of oligonucleotides containing complementary tag sequences, they may be substituted for

standard sequencing adaptors. Thus, use of DCS does not require any significant deviations from the normal workflow of sample preparation for Illumina DNA sequencing. Moreover, the DCS approach can be generalized to nearly any sequencing platform because a double-stranded SMI tag can be incorporated into other existing adaptors, or for sequencing approaches that do not require adaptors, a double-stranded SMI tag can be ligated onto duplex DNA sample prior to sequencing. The compatibility of DCS with existing sequencing workflows, the potential for greatly reducing the error rate of DNA sequencing, and the multitude of applications for the double-stranded SMI sequences validate DCS as a technique that may play a general role in next generation DNA sequencing.

[0087] The following examples are intended to illustrate various embodiments of the invention. As such, the specific embodiments discussed are not to be construed as limitations on the scope of the invention. It will be apparent to one skilled in the art that various equivalents, changes, and modifications may be made without departing from the scope of invention, and it is understood that such equivalent embodiments are to be included herein. Further, all references cited in the disclosure are hereby incorporated by reference in their entirety, as if fully set forth herein.

EXAMPLES

Example 1: Generation of SMI Adaptor Molecules and their use in sequencing double-stranded target DNA

Materials and Methods

[0088] *Materials.* Oligonucleotides were from IDT and were ordered as PAGE purified. Klenow exo- was from NEB. T4 ligase was from Enzymatics.

[0089] *DNA isolation.* Genomic DNA was isolated from normal human colonic mucosa by sodium iodide extraction (Wako Chemicals USA).

[0090] *Adaptor synthesis.* The adaptors were synthesized from two oligos, designated as:

the primer strand:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT (SEQ ID NO:1); and

the template strand:

/5phos/ACTGNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTG AACTCCAGTCAC (SEQ ID NO:2).

[0091] The two adaptor strands were annealed by combining equimolar amounts of each oligo to a final concentration of 50 micromolar and heating to 95°C for 5 minutes. The oligo mix was allowed to cool to room temperature for over 1 hour. The annealed primer-template complex was extended in a reaction consisting of 40 micromolar primer-template, 25 units Klenow exo-DNA polymerase (New England Biolabs), 250 micromolar each dNTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, and 1 mM dithiothreitol (DTT) for 1 hour at 37°C. The product was isolated by ethanol precipitation. Due to the partial A-tailing property of Klenow exo-, this protocol results in a mixture of blunt-ended adapters and adapters with a single-nucleotide A hverhang. A single-nucleotide A overhang was added to residual blunt fragments by incubating the adapters with 25 units Klenow exo-, 1 mM dATP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM MgCl2, and 1 mM dithiothreitol (DTT) for 1 hour at 37°C. The product was again ethanol precipitated and resuspended to a final concentration of 50 micromolar.

[0092] *Sequencing library preparation.* 3 micrograms of DNA was diluted into 130 microliters of TE buffer (10 mM tris-HCl, pH 8.0, 0.1 M EDTA) and was sheared on the Covaris AFA system with duty cycle 10%, intensity 5, cycles/burst 200, time 20 seconds x 6, temperature 4°C. DNA was purified with 2 volumes of Agencourt AMPure XP beads per the manufacturer's protocol. After end-repair with the NEB end-repair kit per the manufacturer's protocol, DNA fragments larger than the optimal range of ~200-500 bp were removed by adding 0.7 volumes of AMPure XP beads and transferring the supernatant to a separate tube (fragments larger than 500 bp bind to the beads and are discarded). An additional 0.65 volumes of AMPure XP beads were added (this step allows fragments of approximately 200bp or greater to bind to the beads). The beads were washed and DNA eluted. DNA was then T-tailed in a reaction containing 5 units

Klenow exo-, 1 mM dTTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.9, 10 mM MgCl2, 1 mM. The reaction proceeded for 1 hour at 37 C. DNA was purified with 1.2 volumes of AMPure XP beads. The custom adaptors were ligated by combining 750 ng of T-tailed DNA with 250 pmol adaptors in a reaction containing 3000 units T4 DNA ligase, 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 5 mM DTT, 1 mM ATP. The reaction was incubated 25 C for 15 minutes, and purified with 1.2 volumes of AMPure XP beads.

[0093] *Pre-capture amplification.* 375 ng adaptor-ligated DNA was PCR amplified with primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and GTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:4) using the Kappa high-fidelity PCR kit for 8 cycles with an annealing temperature of 60 C. The product was purified with 1.2 volumes of AMPure XP beads.

[0094] *DNA capture.* Target capture was performed with the Agilent SureSelect system per the manufacturer's recommendations, except that capture volumes were performed at one-half of the standard volume. The capture set targeted an arbitrary 758 kb region of the genome consisting of both coding and noncoding sequences. Capture baits were 120 nt in length, and were prepared with the Agilent eArray tool with 3x tiling.

[0095] *Post-capture amplification.* Captured DNA was amplified with PCR primers AATGATACGGCGACCACCGAG (SEQ ID NO:3) and CAAGC AGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGC (SEQ ID NO:5) where XXXXXX indicates the position of a fixed multiplexing barcode sequence). 2 0 fmol of DNA was used per lane for sequencing on an Illumina HiSeq 2000.

[0096] *Data processing.* Reads with intact SMI adaptors include a 12 nucleotide random sequence, followed by a 5 nucleotide fixed sequence. These reads were identified by filtering out reads that lack the expected fixed sequence at positions 13-17. The SMI sequence from both the forward and reverse sequencing reads (i.e., the first and second degenerate n-mer sequences) was computationally added to the read header, and the fixed sequence removed. The first 4 nucleotides located following the adaptor sequence were also removed due to the propensity for ligation and end-repair errors to result in an elevated error rate near the end of the DNA fragments. Reads

having common (i.e., identical) SMI sequences were grouped together, and were collapsed to generate a consensus read. Sequencing positions were discounted if the consensus group covering that position consisted of fewer than 3 members, or if fewer than 90% of the sequences at that position in the consensus group had the identical sequence. Reads were aligned to the human genome with the Burrows-Wheeler Aligner (BWA). The consensus sequences were then paired with their strand-mate by grouping each 24 nucleotide tag of form AB in read 1 with its corresponding tag of form BA in read 2. Resultant sequence positions were considered only when information from both DNA strands was in perfect agreement. An overview of the data processing workflow is as follows:

1. Discard reads that do not have the 5 nt fixed reference (or "spacer") sequence (CAGTA; SEQ ID NO:6) present after 12 random nucleotides.
2. Combine the 12 nt SMI tags from read 1 and read 2, and transfer the combined 24 nt SMI sequence into the read header.
3. Discard SMIs with inadequate complexity (i.e., those with > 10 consecutive identical nucleotides).
4. Remove the 5 nt fixed reference sequence.
5. Trim an additional 4 nt from the 5' ends of each read pair (sites of error prone end repair).
6. Group together reads which have identical 24 nt SMIs.
7. Collapse to SMI consensus reads, scoring only positions with 3 or more SMI duplicates and >90% sequence identity among the duplicates.
8. For each read in read 1 file having SMI of format AB, group with corresponding DCS partner in read 2 with SMI of format BA.
9. Only score positions with identical sequence among both DCS partners.
10. Align reads to the human genome.

[0097] Code for carrying out the workflow may be pre-existing or may involve programming within the skill of those in the art. In some embodiments, however, the Python code, which is illustrated in Figure 10, may be used for carrying out the pairing

and scoring of partner strands according to steps 8 and 9 of the workflow described above.

Overview

[0098] To overcome limitations in the sensitivity of variant detection by single-stranded next-generation DNA sequencing, an alternative approach to library preparation and analysis was designed, which is known herein as Duplex Consensus Sequencing (DCS) (Figure 1). The DCS method described herein involves tagging both strands of duplex DNA with a random, yet complementary double-stranded nucleotide sequence, which is known herein as a double-stranded single molecule identifier (SMI) sequence. The SMI sequences (in this case, double stranded SMI sequences) are incorporated into the SMI adaptor molecules by introducing a single-stranded randomized nucleotide sequence into one adapter strand and the extending the opposite strand with a DNA polysmerase to yield a complementary, double-stranded SMI sequence (Figure 2). The individually tagged strands are then PCR amplified. Every duplicate that arises from a single strand of DNA will have the same SMI, and thus each strand in a DNA duplex pair generates a distinct, yet related population of PCR duplicates after amplification owing to the complementary nature of the SMIs on the two strands of the duplex. Comparing the sequence obtained from each of the two strands comprising a single molecule of duplex DNA facilitates differentiation of sequencing errors from true mutations. When an apparent mutation is, due to a PCR or sequencing error, the substitution will only be seen on a single strand. In contrast, with a true DNA mutation, complementary substitutions will be present on both strands (see Figure 4C).

[0099] Following tagging with a double-stranded SMI and PCR amplification, a family of molecules is obtained that arose from a single DNA molecule; members of the same PCR "family" are then grouped together by virtue of having a common (i.e., the same) SMI tag sequence. The sequences of uniquely tagged PCR duplicates are subsequently compared in order to create a PCR consensus sequence. Only DNA positions that yield the same DNA sequence in a specified proportion of the PCR duplicates in a family, such as 90% of the duplicates in one embodiment, are used to create the PCR consensus sequence. This step filters out random errors introduced

during sequencing or PCR to yield the PCR consensus sequences, each of which derives from an individual molecule of single-stranded DNA. This set of PCR consensus sequences are called single strand consensus sequences (SSCSs).

[00100] Next, PCR consensus sequences arising from two complementary strands of duplex DNA can be identified by virtue of the complementary SMIs (Figure 3) to identify the "partner SMI." Specifically, a 24-nucleotide SMI consists of two 12-nucleotide sequences that can be designated XY. For an SMI of form XY in read 1, the partner SMI will be of form YX in read 2. An example to illustrate this point is given in Figure 4. Following partnering of two strands by virtue of their complementary SMIs, the sequences of the strands are compared. Sequence reads at a given position are kept only if the read data from each of the two paired strands is in agreement.

Results

[00101] In order to label or tag each of the strands of duplex DNA with unique complementary tags, adaptors which contain the standard sequences required for the Illumina HiSeq system were synthesized, but with addition of a double-stranded, complementary SMI sequence (or "tag") of 12 random nucleotides (or a random "degenerate sequence") per strand. Target DNA molecules having a random SMI sequence n-mer that is 12 nucleotides in length on each end will therefore have a unique 24 nucleotide SMI sequence. The adaptors were prepared (Figure 2) from two partially complementary oligonucleotides, one of which has a single-stranded 12 nucleotide random nucleotide sequence (i.e. a first random degenerate nucleotide n-mer sequence) followed by a single stranded fixed reference sequence that is 4 nucleotides in length. The single-stranded random nucleotide tag was converted to a double-stranded, complementary SMI tag by extension with Klenow exo- DNA polymerase and the extended adaptor was purified by ethanol precipitation. Due to the partial A-tailing property of Klenow exo-, this protocol results in a mixture of blunt-ended adaptors and adaptors with a single-nucleotide A overhang (data not shown). A single-nucleotide A-overhang was added to the residual blunt fragments by incubating the adaptors with Klenow exo- DNA polymerase and a high concentration of dATP (1 mM), and purified the adaptors again by ethanol precipitation.

[00102] DNA for sequencing was sheared and end-repaired by standard methods, with size-selection for fragments in the range of ~200-500 bp by size-selective binding to Ampure XP beads. Standard Illumina library preparation protocols involve ligating A-tailed DNA to T-tailed adaptors. However, because A-tailed adaptors were used, the DNA was T-tailed by incubating the end-repaired DNA with Klenow exo- DNA polymerase and 1 mM dTTP. The adaptor-ligated library was PCR amplified and subjected to SureSelect capture, with targeting of an arbitrary 758 kb portion of the genome (DNA coordinates available upon request). The efficiency of adaptor ligation, PCR amplification, DNA capture, and sequencing were comparable to those seen with standard library preparation methods (data not shown). Although Agilent Sure Select probes are used in this example, any suitable method of DNA selection may be used to capture particular target double-stranded DNA sequences. For example, selection and capture may be accomplished by any selection by hybridization method (e.g., Agilent SureSelect, Primer Extension Capture, exploitation of biotinylated PCR amplicons as bait, Agilent HaloPlex) wherein probes that target the desired double-stranded DNA sequence may be recovered by an in-array capture (using probes immobilized on glass slides) or by affinity using magnetic beads in an in-solution capture. In addition, mitochondrial and some other forms of DNA may be isolated by size selection. Alternatively, in some embodiments, no enrichment is performed.

[00103] This protocol was used to sequence DNA isolated from normal colonic mucosa. Mutations were initially scored without consideration of the SMI sequences. PCR duplicates were filtered out with samtools rmdup, a standard tool which uses the shear points of DNA molecules to identify PCR duplicates, as molecules arising from duplicated DNA will have shared shear points. In order to focus specifically on non-clonal mutations, only those positions in the genome with at least 20x coverage and at which fewer than 5% of reads differed from the hg19 reference sequence were considered. This approach resulted in 70.9 million nucleotides of sequence data and 56,890 mutations, indicating an overall mutation frequency of 8.03×10^{-4}, in accord with the error rate of Illumina next-generation sequencing of ~0.1-1% [32].

[00104] Next, the SMI tags were used to group together PCR duplicates that arose from individual single-stranded DNA molecules and to create a consensus sequence

from the family of duplicates. At least 3 PCR duplicates were required, with at least 90% agreement in sequence among all duplicates, to consider a site for mutations. Scoring the mutation frequency as above, again considering only sites with a minimum of 20x coverage and with <5% of reads differing from reference, resulted in 145 million nucleotides of sequence with 6,508 mutations and an overall mutation frequency of $4.47x10^{-5}$, consistent with prior reports [36]. Notably, far more nucleotides of DNA sequence were obtained in this approach (145 million) than in the standard Illumina sequencing approach (70 million) detailed above which is dependent on use of the shear points of single-ended reads to identify PCR duplicates. The improved sequence coverage arose from use of the SMI to identify PCR duplicates, because identifying PCR duplicates by consideration of uniquely sheared DNA ends is fundamentally limited by the small number of possible shear points that overlap a given position of the genome and the propensity for specific genomic regions to be more readily undergo shearing. Thus filtering PCR duplicates by using shear points resulted in discarding a large portion of the reads.

[00105] Finally, the complementary nature of the double-stranded SMI sequences was used to identify pairs of consensus groups that arose from complementary DNA strands. Sequence reads were considered only when the read data from each of the two strands is in perfect agreement. In a pilot experiment, after grouping of PCR duplicates as above, 29,409 SMI partner pairs were found, indicative that fewer than 1% of tags had their corresponding partner tag present in the library. The low recovery of tag pairs was most likely due to inadequate amplification of the starting DNA library. Among these tag-pairs, 24,772 duplex consensus strands were identified with an average strand length of 82 nucleotides, resulting in 2 million nucleotides of DNA consensus sequence. The sequences of the paired duplex strands disagreed at 3,585 of the nucleotide positions, indicative of single-stranded errors (i.e. PCR or sequencing errors); these sites of disagreement were removed, leaving only bases at which the sequence of both duplex strands were in perfect agreement. Next, as above, analysis of mutation frequencies was restricted to sites with at least 10x coverage and at which fewer than 10% of reads disagreed from the hg19 reference sequence. Because the 2 million nucleotides of read data were spread across a 758kb target, our average depth

was only ~3x. Thus only 14,464 nucleotides of DNA sequence corresponded to sites with at least 10x depth. Among these sites, zero mutations were seen. To increase the number of tag pairs considered, analysis described above was repeated, but PCR duplicates were grouped with a minimum of only 1 duplicate per site. This resulted in 28,359 nucleotides of DNA sequence with at least 10x depth. Again, no mutations were detected.

[00106] Current experiments are being performed on vastly smaller target DNA molecules (ranging from ~300 bp to ~20 kb in size). Use of smaller DNA targets will allow for much greater sequencing depth, and far more accurate assessment of the background mutation rate of the assay. In addition, the protocol has been modified to incorporate a greater number of PCR cycles initiated off a smaller number of genome equivalents, which will increase the fraction of tags for which both of the partner tag strands have been sufficiently amplified to be represented in the final sequence data. Indeed, among the 3.6 million SMIs present in our initial library which underwent PCR duplication, 1.5 million of the SMIs were present only once, indicating insufficient amplification of the DNA due in part to the low number of PCR cycles used.

Example 2: Duplex Sequencing of Human Mitochondrial DNA.

Materials and methods

[00107] In addition to those described in Example 1 above, the following materials and methods were also used.

[00108] *DNA Isolation*. Mitochondrial DNA was isolated as previously described (4).

[00109] *Data processing*. The entire human genome sequence (hg19) was used as reference for the mitochondrial DNA experiment, and reads that mapped to chromosomal DNA were removed. Reads sharing identical tag sequences were then grouped together and collapsed to consensus reads. Sequencing positions were discounted if the consensus group covering that position consisted of fewer than three members or if fewer than 90% of the sequences at that position in the consensus group had the identical sequence. A minimum group size of three was selected because next-generation sequencing systems have an average base calling error rate of ~1/100.

Requiring the same base to be identified in three distinct reads decreases the frequency of single-strand consensus sequence (SSCS) errors arising from base-call errors to $(1/100)3 = 1 \times 10^{-6}$, which is below the frequency of spontaneous PCR errors that fundamentally limit the sensitivity of SSCSs. The requirement for 90% of sequences to agree to score a position is a highly conservative cutoff. For example, with a group size of eight, a single disagreeing read will lead to 87.5% agreement and the position will not be scored. If all groups in an experiment are of size nine or less, this cutoff will thus require perfect agreement at any given position to score the position. Further development of our protocol may allow for less stringent parameters to be used to maximize the number of SSCS and duplex consensus sequence (DCS) reads that can be obtained from a given experiment.

Results

[00110] Having established the methodology for Duplex Sequencing with M13mp2 DNA, which is a substrate for which the mutation frequency and spectrum are fairly well established, it was desired to apply the approach to a human DNA sample. Thus, mitochondrial DNA was isolated from human brain tissue and sequenced the DNA after ligation of Duplex Sequencing adapters. A standard sequencing approach with quality filtering for a Phred score of 30 resulted in a mutation frequency of 2.7 × 10−3, and SSCS analysis yielded a mutation frequency of 1.5 × 10−4. In contrast, DCS analysis revealed a much lower overall mutation frequency of 3.5 × 10−5 (Figure 5A). The frequency of mutations in mitochondrial DNA has previously been difficult to measure directly due in part to sources of error in existing assays that can result in either overestimation or underestimation of the true value. An additional confounder has been that most approaches are limited to interrogation of mutations within a small fraction of the genome [56]. The method of single-molecule PCR, which has been proposed as an accurate method of measuring mitochondrial mutation frequency [56] and is considered resistant to damage-induced background errors [57], has resulted in a reported mitochondrial mutation frequency in human colonic mucosa of 5.9 × 10−5 ± 3.2 × 10−5 [56], which is in excellent agreement with our result. Likewise, mitochondrial DNA sequence divergence rates in human pedigrees are consistent with a mitochondrial mutation frequency of 3–5 × 10−5 [58, 59].

[00111] When the distribution of mutations throughout the mitochondrial genome is considered, the quality filtered reads (analyzed without consideration of the tags) have many artifactual errors, such that identification of mutational hotspots is difficult or impossible (Figure 5B). DCS analysis removed these artifacts (Figure 5C) and revealed striking hypermutability of the region of replication initiation (D loop), which is consistent with prior estimates of mutational patterns in mitochondrial DNA based upon sequence variation at this region within the population [60].

[00112] SSCS analysis produced a strong mutational bias, with a 130-fold excess of G→T relative to C→A mutations (Figure 5D), consistent with oxidative damage of the DNA leading to first-round PCR mutations as a significant source of background error. A high level of oxidative damage is expected in mitochondrial DNA, due to extensive exposure of mitochondria to free radical species generated as a byproduct of metabolism [61]. DCS analysis (Figure 5E) removed the mutational bias and revealed that transition mutations are the predominant replication errors in mitochondrial DNA. The DCS mutation spectrum is in accord with prior estimates of deamination events [62] and T-dGTP mispairing by the mitochondrial DNA polymerase [63] as primary mutational forces in mitochondrial DNA. Furthermore, the mutation spectrum of our mitochondrial data are consistent with previous reports of heteroplasmic mutations in human brain showing an increased load of A→G/T→C and G→A/C→T transitions, relative to transversions [64, 65]. A similar spectral bias has also been reported in mice [62, 66] and in population studies of Drosophila melanogaster [67].

Example 3: Demonstration of error-correction by DCS using randomly sheared DNA ends as Single Molecule Identifiers

Materials and Methods

[00113] In addition to those described in the Examples above, the following materials and methods were also used to demonstrate the capability of DCS analysis to remove sequencing errors

[00114] *Sequencing library preparation.* Genomic DNA was isolated from a derivative of *Saccharomyces cerevisiae* strain SC288 by standard methods. The DNA was randomly sheared by the Covaris AFA system, followed by end-repair, A-tailing,

and ligation of Illumina TruSeq DNA sequencing adaptors, all by standard library preparation methods. The resultant sequence data consisted of an average 32.5 fold depth of the 12 megabase *S. cerevisiae* genome.

[00115] *Data analysis.* The first 10 nucleotides of each sequencing read pair, corresponding to the randomly sheared DNA ends, were combined, such that the first 10 nucleotides of read 1, referred to as A, was combined with the first 10 nucleotides of read 2, referred to as B, to yield an SMI tag of form AB. Reads were grouped according to SMI sequence, and nucleotide reads were considered only if they agreed among at least 90% of family members sharing a given tag sequence. For DCS analysis, a tag of form AB1 is partnered with the corresponding tag of form BA2, and nucleotide positions are considered only when the sequence is in agreement among read pairs with matching tags AB1 and BA2.

Results

[00116] In order to demonstrate the capability of DCS analysis to remove sequencing errors, a sequencing library was prepared under standard conditions with commercially available sequencing adaptors, and the randomly sheared DNA ends were used as SMI's. First, reads were grouped by SMI with a minimum family size of 1 member. Considering only sites with a minimum of 20x coverage and with <5% of reads differing from reference, this analysis resulted in 644.8 million nucleotides of sequence data and 2,381,428 mutations, yielding an overall mutation frequency of 3.69×10^{-3}.

[00117] The data was then subjected to DCS analysis with the SMI tags, searching for tags of form AB1 that have partner tags of form BA2, and considered only positions at which the sequence from the two strands was in perfect agreement. 3.1% of the tags had a matching partner present in the library, resulting in 2.9 million nucleotides of sequence data. The sequences of the duplex strands were not complementary at 40,874 nucleotide positions; these disagreeing positions, representing likely sequencing or PCR errors, were removed from analysis. Again considering positions with at least 20x coverage and <5% of reads differing from reference, 3.0 million nucleotides of sequence data and 157 mutations were obtained, with an overall mutation frequency of 5.33×10^{-5}, indicative of removal of >98% of

mutations seen in raw analysis and thereby demonstrating the capability of DCS to lower the error rate of DNA sequencing.

[00118] To compare this result to the method of Kinde et al. [36], reads were grouped into families by SMI tag as before but filtered for families with a minimum of 3 members. This resulted in 1.4 million nucleotides of sequence data and 61 mutations, with an overall mutation frequency of 4.25×10^{-5}. Thus, the method of Kinde et al., with a minimum family size of 3, resulted in less than half as much resultant sequence data after filtering than was obtained by DCS with a minimum family size of 1. Thus, DCS lowered the error rate of sequencing to a comparable degree to a method considered state-of-the-art, but with less loss of sequencing capacity.

Discussion

[00119] It was demonstrated that DCS analysis, using sheared DNA ends as unique molecular identifiers, results in a lowering of the apparent error rate of DNA sequencing. As this proof-of-concept experiment was performed on a library that was not optimized to maximize recovery of both strands, there were not sufficient strand-pairs recovered to perform DCS analysis with a minimum family size of >1 member. Requiring family sizes >1 is expected to further reduce sequencing errors. Moreover, this analysis was limited in that it did not include ligation of degenerate SMI tag sequences; owing to the limited number of shear points flanking any given nucleotide position, use of shear points as SMIs limits the number of unique molecules that can be sequenced in a single experiment. The use of shear points as SMIs in conjunction with an exogenously ligated SMI tag sequence would allow for increased depth of sequencing at any given nucleotide position.

Example 4: Demonstrations of duplex consensus sequencing

[00120] In addition to those described in Examples 1 and 2 above, the following materials and methods were also used.

Materials and Methods

[00121] *Construction of M13mp2 Variants*. M13mp2 gapped DNA encoding the LacZ α fragment was extended by human DNA polymerase δ [2] and the resultant

products were transformed into Escherichia coli and subjected to blue-white color screening as previously described [3]. Mutant plaques were sequenced to determine the location of the mutation resulting in the color phenotype. A series of mutants, each differing from wild type by a single nucleotide change, were then mixed together with wild-type M13mp2 DNA to result in a single final mixture with distinct mutants represented at ratios of 1/10 (G6267A), 1/100 (T6299C), 1/1,000 (G6343A), and 1/10,000 (A6293T).

[00122] *Oxidative Damage of M13mp2 DNA*. Induction of DNA damage was performed by minor modifications to a published protocol [5]: 300 ng of M13mp2 double-stranded DNA was incubated in 10 mM sodium phosphate buffer, pH 7.0, in the presence of 10 μM iron sulfate and 10 μM freshly diluted hydrogen peroxide. Incubation proceeded for 30 min at 37 °C in open 1.5-mL plastic microcentrifuge tubes.

[00123] *DNA Isolation*. M13mp2 DNA was isolated from E. coli strain MC1061 by Qiagen Miniprep. To allow for greater sequencing depth at a defined region of the M13mp2 genome, an 840-bp fragment was enriched by complete digestion with the restriction enzymes Bsu36I and NaeI (New England Biolabs), followed by isolation of the fragment on an agarose gel by the Recochip system (Takara Bio).

Duplex Consensus Sequencing of M13 DNA removes artifactual sequencing errors.

[00124] The spontaneous mutation rate of M13mp2 DNA has been well established by a number of exquisitely sensitive genetic assays to be 3.0E-6 [53], that is, an average of one spontaneous base substitution error for every 330,000 nucleotides. Thus this substrate is well suited as a control for determining the background error frequency of DNA sequencing. M13mp2 DNA was sheared and ligated to adaptors containing double-stranded complementary SMI sequences by standard protocols, and was subjected to deep sequencing on an Illumina HiSeq 2000 followed by Consensus Sequencing analysis (Figure 6).

[00125] Analysis of the data by standard methods (i.e., without consideration of the double stranded SMI sequences) resulted in an error frequency of 3.8E-03, more than one thousand fold higher than the true mutation frequency of M13mp2 DNA. This

indicates that >99.9% of the apparent mutations identified by standard sequencing are in fact artifactual errors.

[00126] The data were then analyzed by Single Strand Consensus Sequencing (SSCS), using the unique SMI tag affixed to each molecule to group PCR products together in order to create a consensus of all PCR products that came from an individual molecule of single-stranded DNA. This resulted in a mutation frequency of 6.4E-OS, suggesting that -98% of sequencing errors are corrected by SSCS.

[00127] Next, the data were subjected to Duplex Consensus Sequencing (DCS), which further corrects errors by using the complementary SMI tags to compare the DNA sequence arising from both of the two strands of a single molecule of duplex DNA. This approach resulted in a mutation frequency of 2.SE-06, in nearly perfect agreement with the true mutation frequency of M13mp2 DNA of 3.0E-06. The number of nucleotides of DNA sequence obtained by a standard sequencing approach, and after SSCS and DCS analysis, may be found in Table 1 below.

[00128] **Table 1**. Data yield from Duplex Sequencing

	M13mp2 DNA	Mitochondrial DNA
Initial nucleotides	6.5×10^9	6.2×10^9
SSCS nucleotides	8.7×10^7	4.1×10^8
DCS nucleotides	$2.2E \times 10^7$	9.7×10^7
Initial reads per SSCS read	75	15
Initial reads per DCS read	295	64
SSCS reads per DCS read	4	4

Initial nucleotides represent raw reads that contain the expected fixed adapter sequence following 12 degenerate nucleotides and map to the reference genome. Apparent nucleotide loss in converting initial reads to SSCSs occurs because many of the initial reads intentionally represent identical PCR duplicates of single-stranded DNA molecules to allow for removal of sequencing and PCR errors by comparison of the sequence among the duplicates. A minimum of three initial reads are required to produce one SSCS; however, a greater average number is necessary to ensure that most DNA fragments have at least this number of duplicates. Under fully optimized conditions, each DCS read would arise from exactly two SSCS reads (one arising from each strand of the initial molecule of duplex DNA). An SSCS:DCS ratio greater than 2 indicates that the strand partner of some SSCSs was not recovered.

[00129] For an artifactual error to be scored by DCS, complementary artifactual errors must occur on both strands of a molecule of duplex DNA. Thus the background (artifactual) error frequency of DCS may be calculated as: (probability of error on one strand) * (probability of error on other strand) * (probability that both errors are complementary).

[00130] As the background error frequency of SSCS in this experiment was -6E-S, the background error frequency of DCS can be calculated as 6E-S * 6E-S * (1/3) = 1.2E-9. This represents a greater than 3 million fold improvement over the error rate of 3.SE-03 that was obtained by a standard sequencing approach.

Consensus sequencing reveals likely sites of DNA damage

[00131] M13mp2 DNA was sequenced as detailed above, with DCS adaptors containing double-stranded complementary SMIs. The spectrum of mutations obtained with SSCS was determined. Data was filtered to consist of forward-mapping reads from Read 1, i.e. sequencing of the reference strand, and reverse-mapping reads from Read 1, i.e. sequencing of the anti-reference strand. True mutations would result in an equal balance between mutations on the reference strand and their complementary mutation on the anti-reference strand.

[00132] However, SSCS analysis revealed a large number of single-stranded G→T mutations on reads mapping in the forward orientation to the reference genome, with a much smaller number of C→A mutations mapping to the reverse orientation. The spectrum of mutations identified by both SSCS and DCS analysis were examined relative to literature reference values [53] for the M13mp2 substrate (Figure 7A). SSCS analysis revealed a large excess of G→A/C→T and G→T/C→A mutations relative to reference ($P < 10^{-6}$, two-sample t test). In contrast, DCS analysis was in excellent agreement with the literature values with the exception of a decrease relative to reference of these same mutational events: G→A/C→T and G→T/C→A ($P < 0.01$). To probe the potential cause of these spectrum deviations, the SSCS data were filtered to consist of forward-mapping reads from read 1 (i.e., direct sequencing of the reference strand) and the reverse complement of reverse-mapping reads from read 1 (i.e., direct sequencing of the antireference strand.) True double-stranded mutations should result in an equal balance of complementary mutations observed on the reference and

antireference strand. However, SSCS analysis revealed a large number of single-stranded G→T mutations, with a much smaller number of C→A mutations (Figure 7B). A similar bias was seen with a large excess of C→T mutations relative to G→A mutations.

[00133] Base-specific mutagenic DNA damage is a likely explanation of these imbalances. Excess G→T mutations are consistent with the oxidative product 8-oxo-guanine (8-oxo-G) causing first round PCR errors and artifactual G→T mutations. DNA polymerases, including those commonly used in PCR, have a strong tendency to insert adenine opposite 8-oxo-G [45, 54], and misinsertion of A opposite 8-oxo-G would result in erroneous scoring of a G→T mutation. Likewise, the excess C→T mutations are consistent with spontaneous deamination of cytosine to uracil [47], a particularly common DNA damage event that results in insertion during PCR of adenine opposite uracil and erroneous scoring of a C→T mutation.

[00134] To determine whether the excess G→T mutations seen in SSCSs might reflect oxidative DNA damage at guanine nucleotides, before sequencing library preparation M13mp2 DNA was incubated with the free radical generator hydrogen peroxide in the presence of iron, a protocol that induces DNA damage [55]. This treatment resulted in a substantial further increase in G→T mutations by SSCS analysis (Figure 8A), consistent with PCR errors at sites of DNA damage as the likely mechanism of this biased mutation spectrum. In contrast, induction of oxidative damage did not alter the mutation spectrum seen with DCS analysis (Figure 8B), indicating that duplex consensus sequences are not similarly susceptible to DNA damage artifacts.

[00135] Furthermore, relative to the literature reference values, DCS analysis results in a lower frequency of G→T/C→A and C→T/G→A mutations (Figure 7A), which are the same mutations elevated in SSCS analysis as a probable result of DNA damage. Notably, the M13mp2 LacZ assay, from which reference values have been derived, is dependent upon bacterial replication of a single molecule of M13mp2 DNA. Thus, the presence of oxidative damage within this substrate could cause an analogous first-round replication error by Escherichia coli, converting a single-stranded damage event into a fixed, double-stranded mutation during replication. The slight reduction in the frequency of these two types of mutations measured by DCS analysis may,

therefore, reflect the absence of damage-induced errors that are scored by the in vivo LacZ assay.

Consensus Sequencing accurately recovers spiked-in control mutations.

[00136] A series of M13mp2 variants were constructed which contain known single base substitutions. These variants were then mixed together at known ratios, and the mixture was prepared for sequencing with DCS adaptors with double-stranded complementary SMIs and was sequenced on an Illumina HiSeq 2000. The data was then analyzed by consensus sequencing (Figure 9). With conventional analysis of the data (i.e. without consideration of the SMI tags), variants present at a level of <1/100 could not be accurately identified. This limitation occurs because at any given position, artifactual mutations are seen at a level of nearly 1/100.

[00137] In contrast, when the data is analyzed by Single Strand Consensus Sequencing (SSCS) with ~20,000 fold depth, accurate recovery of mutant sequence is seen down to one mutant molecule per 10,000 wild type molecules. Duplex Consensus Sequencing (DCS), which was not performed on this sample, would allow for detection of even rarer mutations.

REFERENCES

The references, patents and published patent applications listed below, and all references cited in the specification above are hereby incorporated by reference in their entirety, as if fully set forth herein.

[1] Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31-46.
[2] Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135-45.
[3] Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, et al. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc Natl Acad Sci USA. 2011;108:13177-82.
[4] Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature. 2011;480:368-71.
[5] García-Garcerà M, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, et al. Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. PLoS ONE. 2011;6:e24161.
[6] Fordyce SL, Ávila-Arcos MC, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen FT, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. BioTechniques. 2011;51:127-33.
[7] Druley TE, Vallania FLM, Wegner DJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA. Nat Methods. 2009;6:263-5.
[8] Out AA, van Minderhout IJHM, Goeman JJ, Ariyurek Y, Ossowski S, Schneeberger K, et al. Deep sequencing to reveal new variants in pooled DNA samples. Hum Mutat. 2009;30:1703-12.

[9] Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA. 2008;105:16266-71.
[10] Chiu RWK, Akolekar R, Zheng YWL, Leung TY, Sun H, Chan KCA, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011;342:c7401.
[11] Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008;105:10513-8.
[12] Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2105;481:506-9.
[13] Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. Science Translational Medicine. 2009;1:12ra23-12ra23.
[14] Hyman RW, Herndon CN, Jiang H, Palm C, Fukushima M, Bernstein D, et al. The dynamics of the vaginal microbiome during infertility therapy with in vitro fertilization-embryo transfer. J Assist Reprod Genet. 2012;29:105-15.
[15] LaTuga MS, Ellis JC, Cotton CM, Goldberg RN, Wynn JL, Jackson RB, et al. Beyond bacteria: a study of the enteric microbial consortium in extremely low birth weight infants. PLoS ONE. 2011;6:e27858.
[16] Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: interindividual variation and dynamic response to diet. Genome Res. 2011;21:1616-25.
[17] Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011;49:3463-9.
[18] Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS ONE. 2011;6:e24907.

[19] Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci USA. 2008;105:13081-6.
[20] De Grassi A, Segala C, Iannelli F, Volorio S, Bertario L, Radice P, et al. Ultradeep Sequencing of a Human Ultraconserved Region Reveals Somatic and Constitutional Genomic Instability. PLoS Biol. 2010;8:e1000275.
[21] Zagordi O, Klein R, Däumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Research. 2010;38:7400-9.
[22] Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17:1195-201.
[23] Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, et al. Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat Methods. 2012;9:78-80.
[24] He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464:610-4.
[25] Ameur A, Stewart JB, Freyer C, Hagström E, Ingman M, Larsson N-G, et al. Ultra-Deep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. PLoS Genet. 2011;7:e1002028.
[26] Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng. 2003;96:317-23.
[27] Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. Nucleic Acids Research. 1990;18:1687-91.
[28] Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 2008;5:1005-10.
[29] Salk J, Fox E, Loeb L. Mutational heterogeneity in human cancers: origin and consequences. Annual Review of Pathology. 2009;5:51-75.

[30] Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods. 2009;6:291-5.
[31] Vandenbroucke I, Van Marck H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques. 2011;51:167-77.
[32] Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, et al. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic Acids Research. 2012;40:e2-e.
[33] Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20:273-80.
[34] Miner BE, Stöger RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Research. 2004;32:e135.
[35] McCloskey ML, Stöger R, Hansen RS, Laird CD. Encoding PCR products with batch-stamps and barcodes. Biochem Genet. 2007;45:761-7.
[36] Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci USA. 2011;108:9530-5.
[37] Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA. 2011;108:20166-71.
[38] Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011;9:72-4.
[39] Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. Nucleic Acids Research. 2011;39:e81-e.

[40] Shiroguchi K, Jia TZ, Sims PA, Xie XS. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. Proc Natl Acad Sci USA. 2012;109:1347-52.
[41] Fu GK, Hu J, Wang P-H, Fodor SPA. Counting individual DNA molecules by the stochastic attachment of diverse labels. Proc Natl Acad Sci USA. 2011;108:9026-31.
[42] Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. Embryonic stem cells and somatic cells differ in mutation frequency and type. Proc Natl Acad Sci USA. 2002;99:3586-90.
[43] Lindahl T, Wood RD. Quality control by DNA repair. Science. 1999;286:1897-1905.
[44] Kunkel, TA. Mutational specificity of depurination. Proc Natl Acad Sci USA. 1984;81:1494-98.
[45] Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature. 1991;349:431-4.
[46] Stiller M, Green RE, Ronan M, Simons JF, Du L, He W., et al. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Acad Sci USA. 2006;103:13578-84.
[47] Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obsbet Gynecol. 2011;204:205e1-11.
[48] Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch M, D'Ascenzo M, et al. Whole exome capture in solution with 3 Gbp of data. Genome Biol. 2010;11:R62:1-8.
[49] Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010; 38:159e1---8.
[50] Kaur M, Makrigiorgos GM. Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. Nucleic Acids Res. 2003; 31:26e1---7.
[51] Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature 461, 814–818.

[52] Lynch AM, Sasaki jC, Elespuru R, Jacobson-Kram D, Thybaud V, et al. New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen. 2011; 52(3):205-23.
[53] Thomas DC, Roberts JD, Sabatino RD, Myers TW, et al. Fidelity of mammalian DNA replication and replicative DNA polymerases. Biochemistry. 1991; 30:11751-9.
[54] Kasai H, et al.(1993) Formation, inhibition of formation, and repair of oxidative 8-hydroxyguanine DNA damage. Basic Life Sci 61:257–262.
[55] McBride TJ, Preston BD, Loeb LA (1991) Mutagenic spectrum resulting from DNA damage by oxygen radicals. Biochemistry 30:207–213.
[56] Greaves LC, et al. (2009) Quantification of mitochondrial DNA mutation load. Aging Cell 8:566–572.
[57] Kraytsberg Y, Nicholas A, Caro P, Khrapko K (2008) Single molecule PCR in mtDNA mutational analysis: Genuine mutations vs. damage bypass-derived artifacts. Methods 46:269–273.
[58] Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501–509.
[59] Parsons TJ, et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368.
[60] Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032.
[61] Kennedy SR, Loeb LA, Herr AJ (2011) Somatic mutations in aging, cancer and neurodegeneration. Mech Ageing Dev.
[62] Vermulst M, et al. (2007) Mitochondrial point mutations do not limit the natural lifespan of mice. Nat Genet 39:540–543.
[63] Song S, et al. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. Proc Natl Acad Sci USA 102:4990–4995.
[64] Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Hum Mol Genet 11:133–145.

[65] Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc Natl Acad Sci USA 93:12382–12387.

[66] Khaidakov M, Heflich RH, Manjanatha MG, Myers MB, Aidoo A (2003) Accumulation of point mutations in mitochondrial DNA of aging mice. Mutat Res 526:1–7.

[67] Haag-Liautard C, et al. (2008) Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Biol 6:e204.

Doc Code: CRF.TRNS.REQ
Document Description: Request for Transfer of a Computer Readable Form

PTO/SB/93 (11-14)
Approved for use through 07/31/2019. OMB 0651-0024
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.

REQUEST FOR TRANSFER OF A COMPUTER READABLE FORM UNDER 37 CFR 1.821(e)

Application No.:	16/118,306	First Named Inventor:	Michael SCHMITT
Filing Date:	2018-08-30	Practitioner Docket No.:	72227-8043.US07
Title of the Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

The sequence information in the paper copy or PDF file of the Sequence Listing submitted in this application in accordance with 37 CFR 1.821(c),

- ☒ as part of the originally-filed specification of this application
- ☐ as a separate amendment of the specification filed
  - ☐ herewith
  - ☐ on _____

is identical to the sequence information in the

- ☒ only computer readable form
- ☐ last filed computer readable form
- ☐ other computer readable form, which was filed on _____

in application number 15/660,785, which was filed on 07/26/2017.

This computer readable form was compliant with 37 CFR 1.821-1.825, and applicant hereby requests that it be used as the computer readable form for the present application, in accordance with 37 CFR 1.821(e).

The above referenced paper copy or PDF file of the Sequence Listing contains no new matter.

A sequence listing text file submitted via EFS-Web that complies with the requirements of 37 CFR 1.824(a) (2)-(6) and (b) (i.e., is a compliant sequence listing ASCII text file), serves as both the paper copy required by 37 CFR 1.821(c) and the CRF required by 37 CFR 1.821(e). If a user submits a compliant sequence listing ASCII text file via EFS-Web, the U.S. Patent and Trademark Office will not carry out a request to use a compliant computer readable "Sequence Listing" that is already on file for another application pursuant to 37 CFR 1.821(e) but will use the sequence listing submitted with the application as originally filed via EFS-Web.

It is understood that upon the transfer of a copy of the computer readable form to this application, the U.S. Patent and Trademark Office will update the copy of the computer readable form to reflect the application number and filing date for this application.

Signature /Lara J. Dueppen/	Date 2018-12-10
Name (Print/Typed) Lara J. Dueppen	Practitioner Registration Number 65002

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

☐ *Total of _____ forms are submitted.

This collection of information is required by 35 U.S.C. 22 and 37 CFR 1.821-1.825. 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 1.11 and 1.14. This collection is estimated to take 6 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.

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.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	**METHODS FOR LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING**

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 14/386,800 filed on 2014-09-20 (national phase of PCT/US2013/032665 filed 2013-03-15).

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I am aware of and acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to me to be material to patentability as defined in 37 CFR 1.56.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

LEGAL NAME OF INVENTOR: **Michael Schmitt**

Signature: _____ **Date:** Nov 6, 2014

PERKINSCOIE

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	**METHODS FOR LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING**

As the below named inventor, I hereby declare that:

This declaration is directed to:

- ☐ The attached application, or
- ☒ United States application or PCT international application number 14/386,800 filed on 2014-09-20 (national phase of PCT/US2013/032665 filed 2013-03-15).

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I am aware of and acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to me to be material to patentability as defined in 37 CFR 1.56.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

LEGAL NAME OF INVENTOR: **Jesse Salk**

Signature: _____ **Date:** 11/5/2014

PERKINSCOIE

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	**METHODS FOR LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING**

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 14/386,800 filed on 2014-09-20 (national phase of PCT/US2013/032665 filed 2013-03-15).

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I am aware of and acknowledge the duty to disclose to the U.S. Patent and Trademark Office all information known to me to be material to patentability as defined in 37 CFR 1.56.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

LEGAL NAME OF INVENTOR: **Lawrence A. Loeb**

Signature: D. Lawrence Loeb **Date:** 1/6/2014

PERKINSCOIE

Electronic Patent Application Fee Transmittal

Application Number:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Michael Schmitt
Filer:	Lara J. Dueppen/Amy Candeloro
Attorney Docket Number:	072227-8043.US07

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY FILING FEE (ELECTRONIC FILING)	4011	1	75	75
UTILITY SEARCH FEE	2111	1	330	330
UTILITY EXAMINATION FEE	2311	1	380	380
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	2202	10	50	500
Miscellaneous-Filing:				
LATE FILING FEE FOR OATH OR DECLARATION	2051	1	80	80

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**1365**

Electronic Acknowledgement Receipt

EFS ID:	34542018
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Michael Schmitt
Customer Number:	94991
Filer:	Lara J. Dueppen/Amy Candeloro
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	10-DEC-2018
Filing Date:	30-AUG-2018
Time Stamp:	21:10:50
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$1365
RAM confirmation Number	121118INTEFSW21115800
Deposit Account	500665
Authorized User	Amy Candeloro

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

37 CFR 1.17 (Patent application and reexamination processing 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	Applicant Response to Pre-Exam Formalities Notice	2018-12-10_MIP_Response_722278043US7.pdf	65933 52686fb80f948bf91da28cd15c9efefcd238d04d	no	1
Warnings:					
Information:					
2		2018-12-10_Prelim_Amendment_722278043US7.pdf	113401 3d40c4aba92ff9be9ffa1a221cd156bba96ca086	yes	10

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Applicant Arguments/Remarks Made in an Amendment	10	10
	Claims	3	9
	Specification	2	2
	Preliminary Amendment	1	1

Warnings:					
Information:					
3	Applicant Arguments/Remarks Made in an Amendment	2018-12-10_Substitute_Spec_MARKUP_722278043US7.pdf	2667647 221452cf36334ba168222f10c64908a1b3c4bd5e	no	54
Warnings:					
Information:					
4		2018-12-10_Substitute_Spec_CLEAN_722278043US7.pdf	1580255 74a74e297e511470203db957c84cc33ba5bf0fe8	yes	54

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**

	Abstract		54	54	
	Specification		1	53	
Warnings:					
Information:					
5	CRF Statement Paper and CRF are the same	2018-12-10_CRF_Transfer_Request_722278043US7.pdf	232003 3a373e6b0a1bd4e89590a5d59367989c9d322f00	no	2
Warnings:					
Information:					
6	Oath or Declaration filed	2018-12-10_SCHMITT-Declaration_722278043US7.PDF	706014 572ff487b9dc3e3c8bd181bf75f32aabc4d44f9c	no	1
Warnings:					
Information:					
7	Oath or Declaration filed	2018-12-10_SALK-Declaration_722278043US7.PDF	765228 08f551bfb353e556423ca9a6151f3171c59355e6	no	1
Warnings:					
Information:					
8	Oath or Declaration filed	2018-12-10_LOEB-Declaration_722278043US7.PDF	707189 54bb5a96d98a6107bf81c1d8e2b17973505082b3	no	1
Warnings:					
Information:					
9	Fee Worksheet (SB06)	fee-info.pdf	38696 742cc0bbbb514752f66d419eb1a7a47800cb0312	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	6876366		

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.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: MICHAEL SCHMITT, ET AL.	CONF. NO.: 1114
APPLICATION NO.: 16/118,306	ART UNIT: 1636
FILED: AUGUST 30, 2018	EXAMINER:
FOR: METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	

RESPONSE TO NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

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

Dear Commissioner:

In response to the Notice mailed October 10, 2018, enclosed herewith are:

1. a substitute specification in marked and clean versions in compliance with 37 CFR 1.52, 1.121 (b)(3), and 1.125;
2. a Request for Transfer of a Computer Readable Form under 37 CFR 1.821(e);
3. a Preliminary Amendment; and
4. copies of the executed inventor declarations from parent application number 14/386,800.

The statutory basic filing fee, the application search fee, the application examination fee, additional claim fees, and the surcharge, totaling $1365.00 (small entity) are paid herewith. Any deficiency or overpayment is authorized to be charged or credited to our Deposit Account No. 50-0665.

Dated: December 10, 2018

Respectfully submitted,

PERKINS COIE LLP

Customer No. 94991
UW/Perkins Coie LLP
Patent - LA
P.O. Box 1247
Seattle, WA 98111-1247
Phone: (310) 788-3349
Fax: (206) 332-7198

By: /Lara J. Dueppen/
Lara J. Dueppen
Registration No. 65,002

REMARKS

Entry of the substitute specification filed herewith is respectfully requested.

A marked version of the substitute specification is enclosed showing amendments to the specification relative to the PCT specification as filed with the application on August 30, 2018. Formatting changes including adjustments to the line spacing are not shown in markings.

Statement under 37 CFR 1.125

The substitute specification filed herewith under 37 CFR 1.125 contains no new matter.

Claims 1-36 are cancelled in this amendment without any prejudice or disclaimer. Applicant reserves the right to pursue the subject matter of any canceled claims in the present application or in one or more continuing applications. Claims 37-66 are newly added with support throughout the original specification and claims. Because no new matter is introduced, entry and examination of the amended claims presented herewith are respectfully requested.

Dated: December 10, 2018

Respectfully submitted,

Customer No. 94991
UW/Perkins Coie LLP
Patent - LA
P.O. Box 1247
Seattle, WA 98111-1247
Phone: (310) 788-3349
Fax: (206) 332-7198

PERKINS COIE LLP

By: /Lara J. Dueppen/
Lara J. Dueppen
Registration No. 65,002

IN THE CLAIMS

The following complete listing of claims replaces all previous listings of claims in the application.

1-36. (Canceled)

37. (New) A method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising:

amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products;

sequencing the amplified target nucleic acid products generated from each original strand;

confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands;

comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand; and

generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary.

38. (New) The method of claim 37, wherein prior to amplifying each strand, the method further comprises ligating the double-stranded target nucleic acid molecule to an adapter molecule resulting in an asymmetric adapter-target nucleic acid complex such that each original strand generates a distinct yet related set of amplified target nucleic acid products during the amplification step.

39. (New) The method of claim 38, wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

40. (New) The method of claim 39, wherein the adapter molecule further comprises a nucleotide sequence that uniquely tags each original strand of the double-stranded target nucleic acid molecule such that each original strand is distinctly identifiable relative to its original complementary strand.

41. (New) The method of claim 37, wherein the confirming step includes (i) grouping the sequence reads of the amplified target nucleic acid products into families of paired target nucleic acid strand sequences using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the double-stranded target nucleic acid molecule as an identifier, and (ii) separating each individual family of paired target nucleic acid strand sequences into a set of sequence reads from one original strand of the double-stranded target nucleic acid molecule and a set of sequence reads from the other original strand of the double stranded target nucleic acid molecule.

42. (New) The method of claim 37, wherein the confirming step includes confirming the presence of at least about 3 sequence reads of an amplified target nucleic acid product generated from each of the original strands, and wherein the comparing step includes comparing a first consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from one original strand with a second consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from the other original strand.

43. (New) The method of claim 37, further comprising identifying a mutation occurring at a particular position in the consensus sequence as a true mutation.

44. (New) The method of claim 43, further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation.

45. (New) The method of claim 37, wherein generating a consensus sequence of the double-stranded target nucleic acid molecule further comprises identifying one or more nucleotide positions that disagree between the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand, and discounting nucleotide positions in the consensus sequence that are in disagreement to provide an error-corrected sequence read of the double-stranded target nucleic acid molecule.

46. (New) A method for generating a high accuracy DNA sequencing read, comprising:

sequencing a first strand of a double-stranded DNA (dsDNA) molecule to generate a first strand sequencing read;

sequencing a second strand of the same dsDNA molecule to generate a second strand sequencing read;

grouping the first and second sequence reads by a common feature;

comparing the first strand sequencing read to the second strand sequencing; and

generating a consensus sequence comprising a sequence of nucleotide bases, wherein each nucleotide base is identified at a given position in the consensus sequence when a specific nucleotide is complementary between the first strand sequencing read and the second strand sequencing read.

47. (New) The method of claim 46, wherein generating a consensus sequence comprises:

identifying nucleotide positions where the compared first and second strand sequence reads are non-complementary;

identifying nucleotide positions where the compared first and second strand sequence reads are complementary; and

generating a high accuracy consensus sequence read for the dsDNA molecule that includes only the nucleotide positions where the compared first and second strand sequence reads are complementary.

48. (New) The method of claim 46, further comprising comparing the consensus sequence to a respective reference sequence, and identifying one or more nucleotide sequence variations present in the consensus sequence not present in the reference sequence.

49. (New) The method of claim 48, further comprising identifying a variation occurring at a particular position in the consensus sequence as a true mutation.

50. (New) The method of claim 48, further comprising identifying a variation that occurs at a particular position in only one of the first strand sequence read or the second strand sequence read as a potential artifact.

51. (New) The method of claim 46, wherein the dsDNA molecule is one of a population of dsDNA molecules, and wherein the method further comprises generating a consensus sequence for a plurality of the dsDNA molecules, and wherein for each of one or more of the generated consensus sequences, the method further comprises:

comparing the consensus sequence to a reference sequence; and

identifying one or more variations present in the consensus sequence not present in the reference sequence,

wherein the one or more variations present in the consensus sequence are indicative of a cancer risk, a cancer metabolic state, a mutator phenotype, a carcinogen exposure, a chronic inflammation exposure, an age, a neurodegenerative disease, or a combination thereof in an organism from which the dsDNA molecule is derived.

52. (New) The method of claim 46, wherein the common feature comprises one or more DNA molecule features that distinguish the dsDNA molecule from other dsDNA molecules in a sample.

53. (New) The method of claim 52, wherein the one or more DNA molecule features includes a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the dsDNA molecule, or a combination thereof.

54. (New) The method of claim 46, wherein prior to sequencing either the first strand or the second strand, the method further comprises ligating the dsDNA molecule to at least one adapter molecule to form an adapter-dsDNA complex, and wherein the adapter molecule comprises a single molecule identifier (SMI) sequence that alone or in combination with fragment features of the dsDNA molecule, provides the common feature.

55. (New) A method of generating an error-corrected consensus sequence comprising:

ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules;

for each tagged DNA molecule, generating a set of duplicates of an original forward strand of the tagged DNA molecule and a set of duplicates of an original complement strand of the tagged DNA molecule;

creating a single strand consensus sequence (SSCS) from the set of duplicates of the original forward strand and a single strand consensus sequence

(SSCS) from the set of duplicates of the original complement strand;

comparing the SSCS of the original forward strand to the SSCS of the original complement strand; and

generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward strand and the SSCS of the original complement strand are complimentary.

56. (New) The method of claim 55, wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

57. (New) The method of claim 56, wherein the SMI adapter molecule comprises a single-stranded SMI sequence or a double-stranded SMI sequence.

58. (New) The method of claim 56, wherein the tagged DNA molecule comprises an SMI sequence in each of its original forward strand and original complement strand, and preferably wherein the SMI sequence in one strand comprises a first degenerate or semi-degenerate sequence and in the other strand comprises a second degenerate or semi-degenerate sequence that is complementary to the first such sequence.

59. (New) The method of claim 56, wherein the adapter molecule comprises a Y-shape, a U-shape, or a combination thereof.

60. (New) The method of claim 55, wherein the tagged DNA molecule comprises an adapter molecule at each end.

61. (New) The method of claim 55, wherein the comparing step includes pairing the SSCS of the original forward strand to the SSCS of the original complement strand

using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the duplex DNA molecules as an identifier.

62. (New) The method of claim 55, further comprising sequencing one or more of the duplicates of the original forward strand and one or more of the duplicates of the original complement strand of the tagged DNA molecules using a next-generation sequencing procedure.

63. (New) The method of claim 55, wherein the set of duplicates of the original forward strand and the set of duplicates of the original complement strand includes a minimum of 1 member in each set.

64. (New) The method of claim 55, wherein the set of duplicates of the original forward strand and the set of duplicates of the original complement strand includes a minimum of 3 members in each set.

65. (New) The method of claim 55, wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject, and wherein the method further comprises:

comparing the error-corrected consensus sequence to a respective reference sequence; and

detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

66. (New) The method of claim 65, further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

IN THE SPECIFICATION

Please replace the specification currently on file with the enclosed substitute specification.

RAW SEQUENCE LISTING

Loaded by SCORE, no errors detected.

Application Serial Number: 16118306
Source: OPAP
Date Processed by SCORE: 12/13/2018

ENTERED

SEQUENCE LISTING

<110> UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION

<120> METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

<130> 72227.8043.WO00

<140> US 16/118,306
<141> 2018-08-30

<150> US 15/660,785
<151> 2017-07-26

<150> PCT/US2013/032665
<151> 2013-03-15

<150> US 61/613,413
<151> 2012-03-20

<150> US 61/625,623
<151> 2012-04-17

<150> US 61/625,319
<151> 2012-04-17

<160> 9

<170> PatentIn version 3.5

<210> 1
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 1
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatct
58

<210> 2
<211> 50
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (5)..(16)
<223> n is a, c, g, or t

<400> 2
actgnnnnnn nnnnnnagat cggaagagca cacgtctgaa ctccagtcac
50

<210> 3
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 3
aatgatacgg cgaccaccga g
21

<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 4
gtgactggag ttcagacgtg tgc
23

<210> 5
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (25)..(30)
<223> position of fixed multiplexing barcode sequence

<220>
<221> misc_feature
<222> (25)..(30)
<223> n is a, c, g, or t

<400> 5
caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgc
53

<210> 6
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<400> 6
cagta
5

<210> 7
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (5)..(16)
<223> n is a, c, g, or t

<400> 7
actgnnnnnn nnnnnn
16

<210> 8
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(12)
<223> n is a, c, g, or t

<400> 8
nnnnnnnnnn nncagt
16

<210> 9
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(12)
<223> n is a, c, g, or t

<400> 9
nnnnnnnnnn nncagta
17

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/118,306	08/30/2018	1636	1365	072227-8043.US07	30	2

CONFIRMATION NO. 1114

UPDATED FILING RECEIPT

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

OC000000104567481

Date Mailed: 12/19/2018

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Michael Schmitt, Seattle, WA;
Jesse Salk, Seattle, WA;
Lawrence A. Loeb, Bellevue, WA;

Applicant(s)

UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION, Seattle, WA

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 15/660,785 07/26/2017
which is a CON of 14/386,800 09/20/2014 PAT 9752188
which is a 371 of PCT/US2013/032665 03/15/2013
which claims benefit of 61/613,413 03/20/2012
and claims benefit of 61/625,623 04/17/2012
and claims benefit of 61/625,319 04/17/2012

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 10/09/2018
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/118,306**
Projected Publication Date: 03/28/2019
Non-Publication Request: No
Early Publication Request: No
**** SMALL ENTITY ****
Title

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 16/118,306

APPLICATION AS FILED - PART I

	(Column 1)		(Column 2)	SMALL ENTITY		OR	OTHER THAN SMALL ENTITY	
FOR	NUMBER FILED		NUMBER EXTRA	RATE($)	FEE($)		RATE($)	FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A		N/A	N/A	75		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A		N/A	N/A	330		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A		N/A	N/A	380		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	30	minus 20 =	* 10	x 50 =	500	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	2	minus 3 =	*	x 230 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	1285		TOTAL	

APPLICATION AS AMENDED - PART II

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

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

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

PTO/AIA/80 (07-12)
Approved for use through 11/30/2014. OMB 0651-0035
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.

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

■ Practitioners associated with Customer Number: 94991

OR

☐ Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number

Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignments documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

☐ The address associated with Customer Number:

OR

☐	Firm or Individual Name			
	Address			
	City		State	Zip
	Country			
	Telephone		Email	

Assignee Name and Address: University of Washington through its Center for Commercialization
4311 11th Ave NE, Ste 500, Seattle, WA 98105

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be Filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of The practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee

Signature	Jmeyers	Date	8.6.13
Name	Jessica M. Meyers	Telephone	
Title	Patent Portfolio Manager		

This collection of information is required by 37 CFR 1.31, 1.32 and 1.33. 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 1.11 and 1.14. This collection is estimated to take 3 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.

CENTER FOR COMMERCIALIZATION
UNIVERSITY of WASHINGTON

December 1, 2010

To: Linden Rhoads
Vice Provost, UW Center for Commercialization

From: Jasbir (Jesse) S. Kindra
Director, Intellectual Property Management
UW Center for Commercialization

RE: Signing Authority

In accordance with the letter to me of February 12, 2010 delegating signature authority to me with authority to sub-delegate, I hereby sub-delegate signing authority to Jessica Meyers for the documents listed below:

Small Entity Statements
Power of Attorney
Authorization for Attorney to Perform Work
Assignment Agreements
Non-Disclosure Agreements

The signature authority granted to Jessica Meyers does not include the authority to sub-delegate.

Box 354990 4311 11th Ave. NE, Suite 500, Seattle, WA 98105-4608

206.543.3970 fax 206.543.0586 uwc4c@u.washington.edu depts.washington.edu/uwc4c

Electronic Acknowledgement Receipt

EFS ID:	34891927
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Michael Schmitt
Customer Number:	94991
Filer:	Yang Tang/Jonathan Petrovich
Filer Authorized By:	Yang Tang
Attorney Docket Number:	072227-8043.US07
Receipt Date:	17-JAN-2019
Filing Date:	30-AUG-2018
Time Stamp:	18:49:30
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	Assignee showing of ownership per 37 CFR 3.73	2019-01-17_373c_form_72227-8043US07.PDF	120315 ec13a21efaef3859c5717ec07df2439438fa90da	no	3

Warnings:

Information:					
2	Power of Attorney	UW_C4C_General_POA.PDF	105512 a85c7a7c7956f05876fd4f5af5b10e8a82221012	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	225827		

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/AIA/96 (08-12)
Approved for use through 01/31/2013. 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 displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION

Application No./Patent No.: 16/118,306 Filed/Issue Date: 08/30/2018

Titled: METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION, a university

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. ☑ The assignee of the entire right, title, and interest.

2. ☐ An assignee of less than the entire right, title, and interest (check applicable box):

   ☐ The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   ☐ There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. ☐ The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. ☐ The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. ☑ An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel 038486, Frame 0865, or for which a copy thereof is attached.

B. ☐ A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

   1. From: _____ To: _____

      The document was recorded in the United States Patent and Trademark Office at

      Reel _____, Frame _____, or for which a copy thereof is attached.

   2. From: _____ To: _____

      The document was recorded in the United States Patent and Trademark Office at

      Reel _____, Frame _____, or for which a copy thereof is attached.

This collection of information is required by 37 CFR 3.73(b). 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 1.11 and 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/AIA/96 (08-12)
Approved for use through 01/31/2013. 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 displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☑ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Yang Tang/	2019-01-17
Signature	Date
Yang Tang	55663
Printed or Typed Name	Title or Registration 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.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/118,306	08/30/2018	Michael Schmitt	072227-8043.US07

CONFIRMATION NO. 1114

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

POA ACCEPTANCE LETTER

OC000000105276258

Date Mailed: 01/24/2019

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 01/17/2019.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/dtdinh/

CORRECTED ADS FORM

Application Number	16118306
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Inventor Information

****If no data is shown, no data has been corrected****

	Data of Record	Updated Data
Order Number	1	
Name	~~Michael Schmitt~~	Jesse Salk
Residence Information		
Residency	us-residency	
City	Seattle	
State	WA	
Country of Residence	US	
Mailing Address of Inventor		
Address 1	4311 11th Avenue NE	
Address 2		
City,State/Province, Postal Code	Seattle WA 98105-4608	
Country	US	

	Data of Record	Updated Data
Order Number	2	
Name	~~Jesse Salk~~	Lawrence A. Loeb
Residence Information		
Residency	us-residency	
City	Seattle	Bellevue
State	WA	
Country of Residence	US	
Mailing Address of Inventor		
Address 1	~~4311 11th Avenue NE~~	4311 11th Avenue NE
Address 2		
City,State/Province, Postal Code	Seattle WA 98105-4608	
Country	US	

	Data of Record	Updated Data
Order Number	3	
Name	~~Lawrence A. Loeb~~	Michael Schmitt
Residence Information		
Residency	us-residency	
City	~~Bellevue~~	Seattle
State	WA	
Country of Residence	US	
Mailing Address of Inventor		
Address 1	~~4311 11th Avenue NE~~	4311 11th Avenue NE
Address 2		
City,State/Province, Postal Code	Seattle WA 98105-4608	
Country	US	

Application Information

	Data of Record	Updated Data
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING	
Attorney Docket Number	072227-8043.US07	
Entity Type	Small	

Domestic Benefit/National Stage Information

****If no data is shown, no data has been corrected****

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121,365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing this information in the application data sheet constitutes the specific reference required by 35 U.S. C. 119(e) or 120, and 37 CFR 1.78(a).

		Data of Record	Updated Data
	Prior Application Status		
	Application Number		
	Continuity Type		
	Prior Application Number		
	Filing Date (YYYY-MM-DD)		
	Patent Number		
	Issue Date (YYYY-MM-DD)		

Foreign Priority Information

****If no data is shown, no data has been corrected****

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX) the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

		Data of Record	Updated Data
	Application Number		
	Country		
	Filing Date		
	Access Code		

Applicant Information

****If no data is shown, no data has been corrected****
Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

	Data of Record	Updated Data
Applicant Type		

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is		
Name of the Deceased or Legally Incapacitated Inventor		
Applicant is an Organization		
Name		
Organization Name		
Address 1		
Address 2		
City,State/Province,Postal Code		
Country		
Phone Number		
Fax Number		
Email Address		

Assignee Information including Non-Applicant Assignee Information

****If no data is shown, no data has been corrected****
Providing this information in the application data sheet does not substitute for compliance with any requirement of part 3 of Title 37 of the CFR to have an assignment recorded in the Office

	Data of Record	Updated Data
Order		
Applicant is an Organization		
Name		

Organization Name		

Mailing Address

Address 1		
Address 2		
City,State/Province,Postal Code		
Country		
Phone Number		
Fax Number		
Email Address		

Signature

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b).

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

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

Signature	/Lara J. Dueppen/	Registration Number	65002
First Name	Lara J.	Last Name	Dueppen

Electronic Patent Application Fee Transmittal

Application Number:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	
Filer:	Lara J. Dueppen/Jonathan Petrovich
Attorney Docket Number:	072227-8043.US07

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
PROCESSING FEE, EXCEPT PROV. APPLS.	2830	1	70	70
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**70**

Electronic Acknowledgement Receipt

EFS ID:	35275994
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	28-FEB-2019
Filing Date:	30-AUG-2018
Time Stamp:	19:56:50
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$70
RAM confirmation Number	030119INTEFSW19564900
Deposit Account	500665
Authorized User	Jonathan Petrovich

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	Application Data Sheet to update/ correct info	CorrectedADS.pdf	82102 a096bce252862958ea00c257615e1fd87f387fbd	no	6
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30727 cefab4c3d85c2676a0c2c26c15c9391d77c19e85	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	112829		

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.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/118,306	08/30/2018	1637	1365	072227-8043.US07	30	2

CONFIRMATION NO. 1114

UPDATED FILING RECEIPT

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

OC000000106305498

Date Mailed: 03/05/2019

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Jesse Salk, Seattle, WA;
Lawrence A. Loeb, Bellevue, WA;
Michael Schmitt, Seattle, WA;

Applicant(s)

UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION, Seattle, WA

Power of Attorney: The patent practitioners associated with Customer Number 94991

Domestic Priority data as claimed by applicant

This application is a CON of 15/660,785 07/26/2017
which is a CON of 14/386,800 09/20/2014 PAT 9752188
which is a 371 of PCT/US2013/032665 03/15/2013
which claims benefit of 61/613,413 03/20/2012
and claims benefit of 61/625,623 04/17/2012
and claims benefit of 61/625,319 04/17/2012

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

Projected Publication Date: 03/28/2019

Non-Publication Request: No

Early Publication Request: No

**** SMALL ENTITY ****

Title

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07

CONFIRMATION NO. 1114

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

37 CFR 1.48(f)
ACKNOWLEDGEMENT LETTER

OC000000106305500

Date Mailed: 03/05/2019

NOTICE OF ACCEPTANCE OF REQUEST UNDER 37 CFR 1.48(f)

This is in response to the applicant's request under 37 CFR 1.48(f) submitted on 02/28/2019.

The request under 37 CFR 1.48(f) to correct the inventorship, to correct or update the name of an inventor, or to correct the order of names of joint inventors is accepted.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/mmasfaw/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07

CONFIRMATION NO. 1114

94991
University of Washington/Perkins Coie LLP
PO Box 1247
Seattle, WA 98111-1247

PUBLICATION NOTICE

Title:METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Publication No.US-2019-0093161-A1
Publication Date:03/28/2019

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Public Records Division. The Public Records Division can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Public Records Division, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently https://portal.uspto.gov/pair/PublicPair. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	72227-8043.US07
Confirmation Number	1114

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
	A1	5308751	A	1994-05-03	Ohkawa et al.	
	A2	6251610	B1	2001-06-26	Gupte et al.	
	A3	6958225	B2	2005-10-25	Dong	
	A4	7214490	B2	2007-05-08	Su et al.	
	A5	7267966	B2	2007-09-11	Dong et al.	
	A6	7297778	B2	2007-11-20	Matsuzaki et al.	
	A7	7452699	B2	2008-11-18	Makrigiorgos et al.	
	A8	7459273	B2	2008-12-02	Jones et al.	
	A9	7741463	B2	2010-06-22	Gormley et al.	
	A10	8148068	B2	2012-04-03	Brenner et al.	
	A11	8153375	B2	2012-04-10	Travers, Kevin et al.	
	A12	8715967	B2	2014-05-06	Casbon, James et al.	
	A13	8741606	B2	2014-06-03	Casbon et al.	
	A14	9080210	B2	2015-07-14	Van Eijk, Michael J., et al.	
	A15	9476095	B2	2016-10-25	Vogelstein, Bert et al.	
	A16	9745627	B2	2017-08-29	Van Eijk, Michael J., et al.	
	A17	9862995	B2	2018-01-09	Patel, Abhijit Ajit	
	A18	9898577	B2	2018-02-20	Van Eijk, Michael J., et al.	
	A19	10011871	B2	2018-07-03	Bielas, Jason H.	
	A20	10023907	B2	2018-07-17	Van Eijk, Michael J., et al.	
	A21	10202646	B2	2019-02-12	Fodor, Stephen, 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
	A22	20070128624	A1	2007-06-07	Gormley et al.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	A23	20080167195	A1	2008-07-10	Li et al.	
	A24	20090298075	A1	2009-12-03	Travers et al.	
	A25	20100069263	A1	2010-03-18	Shendure et al.	
	A26	20100222238	A1	2010-09-02	Smith et al.	
	A27	20110301042	A1	2011-12-08	Steinmann et al.	
	A28	20120165202	A1	2012-06-28	Porreca, Gregory et al.	
	A29	20120238738	A1	2012-09-20	Hendrickson et al.	
	A30	20150024950	A1	2015-01-22	Bielas et al.	
	A31	20160319345	A1	2016-11-03	Gnerre et al.	
	A32	20170247687	A1	2017-08-31	Shendure et al.	
	A33	20180023135	A1	2018-01-25	Van Eijk, Michael J., et al.	
	A34	20180363048	A1	2018-12-20	Bielas, Jason H.	
	A35	20180363049	A1	2018-12-20	Bielas, Jason H.	

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]
	B1 **	2006113422	WO	A2	2006-10-26	Ud Gov Health & Human Serv		
	B2 **	2011021102	WO	A2	2011-02-24	POPULATION GENETICS TECHNOLOGIES LTD.		
	B3 **	2012061832	WO	A1	2012-05-10	ILLUMINA, INC.		

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]
	C1 **	Ameur A, Stewart JB, Freyer C, Hagstrom E, Ingman M, Larsson N-G, et al. UltraDeep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. PLoS Genet. 2011;7:e1 002028.	
	C2 **	Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch M, D'Ascenzo M, et al. Whole exome capture in solution with 3 Gbp of data. Genome Bioi. 2010;11: R62:1-8.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C3 **	Boyd SO, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. Science Translational Medicine. 2009;1 :12ra23-12ra23.	
	C4 **	Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci USA. 2008; 105:13081-6.	
	C5 **	Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, et al. Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat Methods. 2012;9:78-80.	
	C6 **	Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. Nucleic Acids Research. 2011 ;39:e81-e.	
	C7 **	Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. Embryonic stem cells and somatic cells differ in mutation frequency and type. Proc Natl Acad Sci USA. 2002;99:3586-90.	
	C8 **	Chiu RWK, Akolekar R, Zheng YWL, Leung TY, Sun H, Chan KCA, et al. Noninvasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011 ;342:c7401.	
	C9 **	De Grassi A, Segala C, Iannelli F, Volorio S, Bertario L, Radice P, et al. Ultradeep Sequencing of a Human Ultraconserved Region Reveals Somatic and Constitutional Genomic Instability. PLoS Biol. 2010;8:e1000275.	
	C10 **	Ding L, Ley T J, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 21 05;481 :506-9.	
	C11 **	Druley TE, Vallania FLM, Wegner OJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA, Nat Methods. 2009;6 :263-5.	
	C12 **	Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obsbet Gynecol. 2011;204:205e1-11.	
	C13 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated October 4, 2018, 6 pages	
	C14 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated June 8, 2018, 5 pages	
	C15 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated August 10, 2017, 7 pages	
	C16 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated May 13, 2016, 6 pages	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C17 **	EUROPEAN PATENT OFFICE, Extended European Search Report for EP13764186.6, dated September 8, 2015, 3 pages	
	C18 **	Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake Sr. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA. 2008;1 05:16266-71.	
	C19 **	Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, et al. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic Acids Research. 2012;40:e2-e.	
	C20 **	Fordyce SL, Avila-Areas MC, Rockenbauer E, B0rsting C, Frank-Hansen R, Petersen FT, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. BioTechniques. 2011;51 :127-33.	
	C21 **	Fu GK, Hu J, Wang P-H, Fodor SPA. Counting individual DNA molecules by the stochastic attachment of diverse labels. Proc Natl Acad Sci USA. 2011;108:9026-31.	
	C22 **	Garcia-Garcera M, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, et al. Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. PLoS ONE. 2011;6:e24161.	
	C23 **	Greaves LC, et al. (2009) Quantification of mitochondrial DNA mutation load. Aging Cell 8:566-572.	
	C24 **	Haag-Liautard C, et al. (2008) Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Bioi 6:e204.	
	C25 **	He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464:610-4.	
	C26 **	Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501-509.	
	C27 **	Hyman RW, Herndon CN, Jiang H, Palm C, Fukushima M, Bernstein D, et al. The dynamics of the vaginal microbiome during infertility therapy with in vitro fertilization embryo transfer. J Assist Reprod Genet. 2012;29:105-15.	
	C28 **	Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA. 2011;108:20166-71.	
	C29 **	Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc Natl Acad Sci USA 93:12382-12387.	
	C30 **	Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng. 2003;96:317-23.	
	C31 **	Kasai H, et al.(1993) Formation, inhibition of formation, and repair of oxidative 8-hydroxyguanine DNA damage. Basic Life Sci 61:257-262.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C32 **	Kaur M, Makrigiorgos GM. Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. Nucleic Acids Res. 2003;31:26e1-7.	
	C33 **	Kennedy SR, Loeb LA, Herr AJ (2012) Somatic mutations in aging, cancer and neurodegeneration. Mech Ageing Dev 133(4):118-26.	
	C34 **	Khaidakov M, Heflich RH, Manjanatha MG, Myers MB, Aidoo A (2003) Accumulation of point mutations in mitochondrial DNA of aging mice. Mutat Res 526:1-7.	
	C35 **	Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci USA. 2011;108:9530-5.	
	C36 **	KIVIOJA et al., "Counting absolute numbers of molecules using unique molecular identifiers," Nature Methods 9, dated January 1, 2012, pages 72-74	
	C37 **	Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner OJ. Amplification free Illumine sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods. 2009; 6:291-5.	
	C38 **	Kraytsberg Y, Nicholas A, Caro P, Khrapko K (2008) Single molecule PCR in mtDNA mutational analysis: Genuine mutations vs. damage bypass-derived artifacts. Methods 46:269-273.	
	C39 **	Kunkel, TA. Mutational specificity of depurination. Proc Natl Acad Sci USA. 1984; 81:1494-98.	
	C40 **	LaTuga MS, Ellis JC, Cotton CM, Goldberg RN, Wynn JL, Jackson RB, et al. Beyond bacteria: a study of the enteric microbial consortium in extremely low birth weight infants. PLoS ONE. 2011; 6:e27858.	
	C41 **	Lecroq B, Lejzerowicz F, Bechar D, Christen R, Esling P, Baerlocher L, et al. Ultradeep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc Natl Acad Sci USA. 2011;108:13177-82.	
	C42 **	Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Hum Mol Genet 11:133-145.	
	C43 **	Lindahl T, Wood RD. Quality control by DNA repair. Science. 1999;286:1897-1905.	
	C44 **	Lynch AM, Sasaki jC, Elespuru R, Jacobson-Kram D, Thybaud V, et al. New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen. 2011;52(3):205-23.	
	C45 **	Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature. 2011 ;480:368-71.	
	C46 **	McBride T J, Preston BD, Loeb LA (1991) Mutagenic spectrum resulting from DNA damage by oxygen radicals. Biochemistry 30:207-213.	
	C47 **	McCloskey ML, Stager R, Hansen RS, Laird CD. Encoding PCR products with batch-stamps and barcodes. Biochem Genet. 2007; 45:761-7.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C48 **	Metzker ML. Sequencing technologies--the next generation. Nat Rev Genet. 2010;11:31-46.	
	C49 **	Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. Nucleic Acids Research. 1990;18:1687-91.	
	C50 **	Miner BE, Stager RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Research. 2004;32:e135.	
	C51 **	Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GO, et al. The human gut virome: interindividual variation and dynamic response to diet. Genome Res. 2011; 21:1616-25.	
	C52 **	Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008; 105:10513-8.	
	C53 **	Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS ONE. 2011;6:e24907.	
	C54 **	Out AA, van Minderhout IJHM, Goeman JJ, Ariyurek Y, Ossowski S, Schneeberger K, et al. Deep sequencing to reveal new variants in pooled DNA samples. Hum Mutat. 2009;30:1703-12.	
	C55 **	Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., Mcinerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature 461, 814-818.	
	C56 **	Parsons T J, et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363-368.	
	C57 **	Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 2008;5:1005-10.	
	C58 **	ROBERTS et al., "Short template amplicon and multiplex megaprimer - enabled relay (STAMMER) sequencing, a simultaneous approach to higher throughput sequence-based typing of polymorphic genes", Immunogenetics (2010), Vol. 62, 253-60	
	C59 **	Salk J, Fox E, Loeb L. Mutational heterogeneity in human cancers: origin and consequences. Annual Review of Pathology. 2009;5:51-75.	
	C60 **	Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20:273-80.	
	C61 **	Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135-45.	
	C62 **	Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature. 1991;349:431-4.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C63 **	SHIROGUCHI et al., “Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes", PNAS 109, dated January 9, 2012, pages 1347-1652	
	C64 **	Song S, et al. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. Proc Natl Acad Sci USA 102:4990-4995.	
	C65 **	Stiller M, Green RE, Ronan M, Simons JF, Du L, HeW., et al. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Aced Sci USA. 2006; 103:13578-84.	
	C66 **	Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029-1032.	
	C67 **	Thomas DC, Roberts JD, Sabatino RD, Myers TW, et al. Fidelity of mammalian DNA replication and replicative DNA polymerases. Biochemistry. 1991;30:11751-9.	
	C68 **	Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010; 38:159e1-8.	
	C69 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Search Report and Written Opinion for PCT/US2013/032665, dated July 9, 2013, 14 pages	
	C70 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Non-Final Office Action for U.S. Patent Application 16/120,072 mailed December 14, 2018. 16 pages.	
	C71 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Non-Final Office Action for U.S. Patent Application No. 16/120,019. Mail Date: November 19, 2018. 16 pages.	
	C72	**UNITED STATES PATENT AND TRADEMARK OFFICE, Notice of Allowance for U.S. Patent Application No. 15/660,785. Mail Date: December 27, 2018. 8 pages.**	
	C73 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Restriction Requirement for U.S. Patent Application No. 16/120,091. Mail Date: November 23, 2018. 7 pages.	
	C74	**UNITED STATES PATENT AND TRADEMARK OFFICE, Notice of Allowance for U.S. Patent Application No. 16/120,091. Mail Date: March 4, 2019. 9 pages.**	
	C75	**UNITED STATES PATENT AND TRADEMARK OFFICE, Final Office Action for U.S. Patent Application No. 16/120,019. Mail Date: June 7, 2019. 17 pages.**	
	C76 **	Vandenbroucke I, Van Marek H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques. 2011;51 :167-77.	
	C77 **	Vermulst M, et al. (2007) Mitochondrial point mutations do not limit the natural lifespan of mice. Nat Genet 39:540-543.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C78 **	Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17:1195-201.	
	C79 **	Wiemann S, Stegemann J, Grothues D, Bosch A, Estivill X, Schwager C, Zimmermann J, Voss H, Ansorge W. (1995) Simultaneous On-Line DNA Sequencing on Both Strands with Two Fluorescent Dyes. Analytical Biochemistry, Vol. 224, pp. 117-121.	
	C80 **	Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbial. 2011 ;49:3463-9.	
	C81 **	Zagordi O, Klein R, Daumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Research. 2010;38:7400-9.	

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.

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Field

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

REFERENCES INCLUDED

Copies of non-patent references C72, C74, and C75 are included. Copies of the foreign and non-patent references marked with a double asterisk are not included, because they were previously cited by or submitted to the Office in prior application number 15/660,785, filed July 26, 2017, or application number 14/386,800, filed September 20, 2014, and relied upon in this application for an earlier filing date under 35 U.S.C. §120. The instant application number 16/118,306 is a continuation of application number 15/660,785.

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2018-06-21
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Electronic Acknowledgement Receipt

EFS ID:	36379896
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	21-JUN-2019
Filing Date:	30-AUG-2018
Time Stamp:	18:29:38
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	Information Disclosure Statement (IDS) Form (SB08)	2019-06-21_Initial-IDS_72227-8043US07.pdf	223328 c91b5744fea1e8d081d86850d892050ded130a3e	no	9

Warnings:

Information:					
This is not an USPTO supplied IDS fillable form					
2	Non Patent Literature	C72-US15660785_NOA_2018-12-27.PDF	10105659 fee3406c7ba16f779be755c9a2fa89bd6abcd0e5	no	8
Warnings:					
Information:					
3	Non Patent Literature	C74-US16120091_NOA_2019-03-04.PDF	11152630 71aa1682bfb34c52af7fcdfc688c9f095fd7638f	no	9
Warnings:					
Information:					
4	Non Patent Literature	C75-US16120019_FOA_2019-06-07.PDF	15329956 d96d097be7398abfab3266181efe7b609ad0d374	no	17
Warnings:					
Information:					
		Total Files Size (in bytes):	36811573		

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.

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1	6582908	B2	2003-06-24	FODOR et al.	
	A2	7406385	B2	2008-07-29	SORENSON	
	A3	8029993	B2	2011-10-04	MIKAWA	
	A4	9249460	B2	2016-02-02	PUSHKAREV et al.	
	A5	9260753	B2	2016-02-16	XIE et al.	
	A6	10287630	B2	2019-05-14	XIE et al.	
	A7	10287631	B2	2019-05-14	SALK 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
	A8	20040209299	A1	2004-10-21	PINTER et al.	
	A9	20070172839	A1	2007-07-26	SMITH et al.	
	A10	20070172873	A1	2007-07-26	BRENNER et al.	
	A11	20110160078	A1	2011-06-30	FODOR et al.	
	A12	20120208724	A1	2012-08-16	STEEMERS et al.	
	A13	20120220494	A1	2012-08-30	SAMUELS et al.	
	A14	20120244525	A1	2012-09-27	HENDRICKSON	
	A15	20140057799	A1	2014-02-27	JOHNSON et al.	
	A16	20140134610	A1	2014-05-15	PHAM et al.	
	A17	20140329698	A1	2014-11-06	BIGNELL et al.	
	A18	20180363051	A1	2018-12-20	SALK et al.	
	A19	20180363052	A1	2018-12-20	SALK et al.	
	A20	20180363053	A1	2018-12-20	SALK et al.	
	A21	20190093161	A1	2019-03-28	SALK et al.	
	A22	20190093162	A1	2019-03-28	SALK et al.	
	A23	20190119748	A1	2019-04-25	SALK et al.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	A24	20190119749	A1	2019-04-25	SALK et al.	
	A25	20190271040	A1	2019-09-05	SALK et al.	
	A26	20190284626	A1	2019-09-19	SALK et al.	
	A27	20190284627	A1	2019-09-19	SALK et al.	
	A28	20190292597	A1	2019-09-26	SALK et al.	
	A29	20190323082	A1	2019-10-24	SALK et al.	
	A30	20190338358	A1	2019-11-07	SALK et al.	
	A31	20190352714	A1	2019-11-21	SALK et al.	

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]
	B1	2012042374	WO	A2	2012-04-05	Taipale et al.		
	B2	2012129363	WO	A2	2012-09-27	Xie et al. (PRESIDENT AND FELLOWS OF HARVARD COLLEGE)		

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]
	C1	BENTLEY, David R. et al., "Accurate whole human genome sequencing using reversible terminator chemistry," Nature 456:53-59 (2008)	
	C2	CHIU, R.W.K. et al., "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma," PNAS 105:20458-20463 (2008)	
	C3	DIEHL, Frank et al., "Basic - Alimentary Tract - Analysis of Mutations in DNA Isolated From Plasma and Stool of Colorectal Cancer Patients", Gastroenterology 135:489–498 (2008)	
	C4	DIEHL, Frank et al., "Detection and quantification of mutations in the plasma of patients with colorectal tumors," PNAS 102:16368-16373 (2005)	
	C5	DING, Li et al., "Analysis of next-generation genomic data in cancer: accomplishments and challenges," Human Molecular Genetics, 19:R188-R196 (2010)	
	C6	EPO, Examination Report for EP13764186.6, dated 1/31/2017, 7 pages.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C7	EWING, B. et al. “Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities” Genome Research (1998) Vol. 8, pp. 186-194.	
	C8	FLEISCHHACKER, M. et al., "Circulating nucleic acids (CNAs) and cancer-A survey," Biochimica et Biophysica Acta. 1775:181-232 (2007)	
	C9	FONG, Siew Lee et al., "Comparison of 7 Methods for Extracting Cell-Free DNA from Serum Samples of Colorectal Cancer Patients," Clinical Chemistry, 55:587-598 (2009)	
	C10	FORSHEW, T. et al. “Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA” Science Translational Medicine (2012) Vol. 4, Issue 136, 136ra68, 14 pages.	
	C11	GORDON, David J. et al., "Causes and consequences of aneuploidy in cancer," Nature Reviews Genetics 13:189-203 (2012)	
	C12	JIANG, Hui et al., "SeqMap: mapping massive amount of oligonucleotides to the genome," Bioinformatics 24:2395-2396 (2008)	
	C13	KAO, Wei-Chun et al, "BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing," Genome Research 19:1884-1895 (2009)	
	C14	KIRCHER, Martin et al., "Improved base calling for the Illumina Genome Analyzer using machine learning strategies," Genome Biology 10:R83 (2009)	
	C15	KIRSCH, Stefan et al., "Sequence error storms and the landscape of mutations in cancer," PNAS 109:14289-14290 (2012)	
	C16	LEDERGERBER, Christian et al., "Base-calling for next-generation sequencing platforms," Briefings in Bioinformatics 12:489-497 (2011)	
	C17	LI, Heng et al., "Fast and accurate short read alignment with Burrows–Wheeler transform," Bioinformatics 25:1754-1760 (2009)	
	C18	LI, Heng et al., "Mapping short DNA sequencing reads and calling variants using mapping quality scores," Genome Research, 18:1851-1858 (2008)	
	C19	LIANG, Kuo-Ching et al., "Bayesian Basecalling for DNA Sequence Analysis Using Hidden Markov Models," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4:430-440 (2007)	
	C20	LIAO, Gary J.W. et al., "Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles," Clinical Chemistry 57:92-101 (2011)	
	C21	LO, Y.M. Dennis et al., "Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis," Am. J. Hum. Genet. 62:768-775 (1998)	
	C22	LUNTER, Gerton et al., "Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads," Genome Research 21:936-939 (2011)	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C23	McKERNAN, Kevin Judd et al., "Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding," Genome Research 19:1527-1541 (2009)	
	C24	MERTES, Florian et al., "Targeted enrichment of genomic DNA regions for next-generation sequencing," Briefings in Functional Genomics, 10:374-386 (2011)	
	C25	MEYER, M. et al. “Targeted high-throughput sequencing of tagged nucleic acid samples” Nucleic Acids Research (2007) Vol. 35, No. 15, e97, 5 pages.	
	C26	MEYERSON, Matthew et al., "Advances in understanding cancer genomes through second-generation sequencing," Nature Reviews Genetics, 11:685-696 (2010)	
	C27	NARAYAN, A. et al. “Ultrasensitive Measurement of Hotspot Mutations in Tumor DNA in Blood Using Error-Suppressed Multiplexed Deep Sequencing” Cancer Research (2012) Vol. 72, No. 14, pp. 3492-3498.	
	C28	NIELSEN, Rasmus et al., "Genotype and SNP calling from next-generation sequencing data," Nature Reviews Genetics, 12:443-451 (2011)	
	C29	QUINLAN, Aaron R. et al., "Pyrobayes: an improved base caller for SNP discovery in pyrosequences," Nature Methods, 5:179-181 (2008)	
	C30	REDON, Richard et al., "Global variation in copy number in the human genome," Nature 444:444-454 (2006)	
	C31	RIZZO, J. et al. “Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing” Cancer Prevention Research (2012), Vol. 5, No. 7, pp. 887-900.	
	C32	SCHMITT, Michael W. et al., "Detection of ultra-rare mutations by next-generation sequencing," PNAS 109:14508-14513 (2012)	
	C33	SCHWARZENBACH, H. et al., "Cell-free nucleic acids as biomarkers in cancer patients," Nature Reviews, 11, 426-437, 2011	
	C34	SCHWEIGER, Michal R. et al., "Genome-Wide Massively Parallel Sequencing of Formaldehyde Fixed-Paraffin Embedded (FFPE) Tumor Tissues for Copy-Number- and Mutation-Analysis," PLOS One 4:e5548 (2009)	
	C35	SEHNERT, Amy J. et al., "Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood," Clinical Chemistry 57:1042-1049 (2011)	
	C36	SPARKS, Andrew B., et al., "Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy," Prenatal Diagnosis, 32:3-9 (2012)	
	C37	TEER, Jamie K. et al, "Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding," Genome Research 20:1420-1431 (2010)	
	C38	USPTO, Application No. 16/120,019, Notice of Allowance, Examiner David C. THOMAS, mailed 10/30/2019. 15 pgs.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C39	USPTO, Application No. 16/120,072, Notice of Allowance, Examiner Stephanie Kane MUMMERT, mailed 6/26/2019. 16 pages.	
	C40	USPTO, Application No. 16/120,091, Second Notice of Allowance, Examiner David C. THOMAS, mailed 6/10/2019. 8 pages.	
	C41	USPTO, Application No. 16/411,045, Non-Final Office Action, Examiner David C. THOMAS, mailed 7/25/2019. 20 pages.	
	C42	USPTO, Application No. 16/411,045, Notice of Allowance, Examiner David C. THOMAS, mail date: 11/15/2019. 16 pgs.	
	C43	USPTO, Application No. 16/411,068, Non-Final Office Action, Examiner David C. THOMAS, mail date: 11/20/2019. 28 pgs.	
	C44	USPTO, Application No. 16/411,069, Non-Final Office Action, Examiner David C. THOMAS, mail date: 11/26/2019. 28 pgs.	
	C45	WAGLE, Nikhil et al., "High-Throughput Detection of Actionable Genomic Alterations in Clinical Tumor Samples by Targeted, Massively Parallel Sequencing," AACR Cancer Discovery 2:82-93 (2011)	

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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2019-12-16
Name/Print	Lara J. Dueppen	Registration Number	65,002

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
5 April 2012 (05.04.2012)

PCT

(10) International Publication Number
WO 2012/042374 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IB2011/002438

(22) International Filing Date:
29 September 2011 (29.09.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1016608.0 1 October 2010 (01.10.2010) GB
1022111.7 30 December 2010 (30.12.2010) GB

(72) Inventors; and
(71) Applicants : **TAIPALE, Anssi Jussi Nikolai** [FI/SE]; Molinvägen 12, 168 50 Bromma (SE). **LINNARSSON, Sten** [SE/SE]; Vindragarvagen 10, S-11750 Stockholm (SE).

(74) Agents: **MEWBURN ELLIS LLP** et al.; 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, 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,

[Continued on next page]

(54) Title: METHOD OF DETERMINING NUMBER OR CONCENTRATION OF MOLECULES

(57) Abstract: A method of determining the number or concentration of molecules in a sample, a method of tracking a substance in a reaction or system, and applications of such methods. Use of nucleic acid as a tracer.

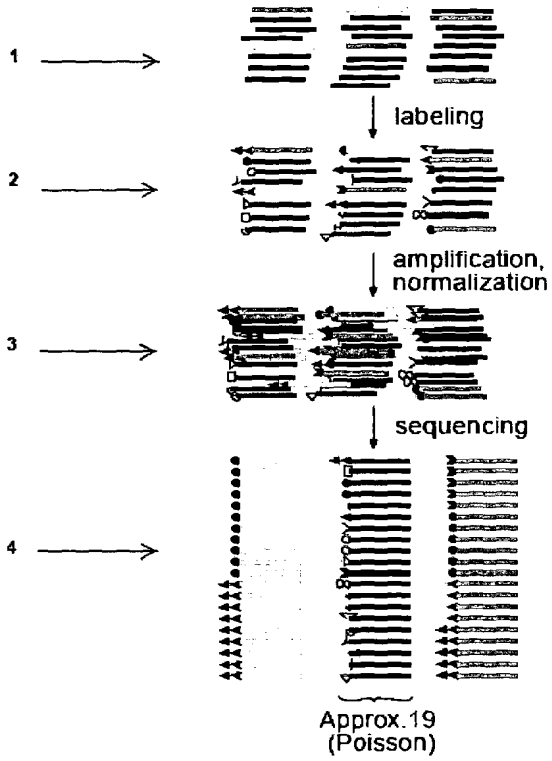


FIGURE 1

WO 2012/042374 A2

NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

METHOD OF DETERMINING NUMBER OR CONCENTRATION OF MOLECULES

The present invention relates to methods for determining the number or concentration of entities in a sample. In particular, the present invention relates to methods for determining the number or concentration of molecules, e.g. biomolecules such as nucleic acid, in a sample.

Advances in molecular biology have made it relatively simple to identify different DNA or RNA species and to copy them. Identification of nucleic acid species can be accomplished by sequencing, i.e. reading the DNA sequence (Maxam and Gilbert, 1977; Sanger and Coulson, 1975); currently millions of molecules can be sequenced in a single day using massively parallel sequencing (Margulies et al,. 2005; Shendure et al., 2005). Efficient copying of DNA-molecules of arbitrary sequence was made possible by molecular cloning (Cohen et al., 1973), and the polymerase chain reaction (Saiki et al,. 1985).

Changes in the relative abundance of a large number of different sequences between two or more samples can in turn be measured using microarray hybridization (Schema et al., 1995) and/or tag sequencing e.g. EST sequencing (Okubo et al., 1992) or sequencing of concatenated mRNA-derived tags (Velculescu et al 1995). However, determining the relative abundance of two different species and/or the absolute number of molecules present in a single sample has proven much more challenging.

Measuring accurate concentrations for multiple different mRNAs simultaneously is difficult because different mRNA species are present in wildly different concentrations in cells. Differences in concentration between high-abundance mRNA and low abundance mRNA can range over six to ten orders of magnitude (see for example Holland et al., 2002). Microarrays have relatively low specificity and linear range, and massively parallel sequencing/molecule counting suffers from the fact that hundreds of thousands of abundant molecules need to be counted before even a single molecule from a rare species is found. Accuracy of measurement requires that a relatively large number of molecules of each species is counted. Measuring a low-abundance mRNA (1 ppm) to an accuracy of ±20% requires counting approximately 10^8 molecules. The range of concentrations can be made more even by subtractive hybridization (Patanjali et al., 1991; Zhulidov et al., 2004), but this results in loss of information about the original concentrations of the mRNA species.

Similarly, methods to amplify the signal by making exact copies of molecules necessarily destroy information about the absolute number of molecules in the original sample, as digital copies derived from two identical molecules are identical. Deriving the absolute number of mRNA or DNA-molecules in the original sample requires precise information about how many

copies of each molecule exist after amplification and/or subtractive hybridization. As efficiency of all known DNA-copying and subtraction processes are dependent on the sequence and length of DNA, determining the number of copies that exist after these procedures is error-prone and not feasible for a large number of different DNA-species. In addition, the cumulative nature of the amplification error makes it very difficult to accurately measure mRNA abundance in samples derived from small amount of starting material. No current method exists to determine absolute number of molecules from complex mixtures efficiently.

It is theoretically possible to count the absolute number of RNA or DNA-molecules by single molecule sequencing (Ozsolak et al., 2010). Using single molecule sequencing to count absolute number of molecules requires a large number of tag counts, as all molecules need to be counted. In addition, as samples from different sources have identical sequences, it is not currently possible to effectively multiplex samples in direct single-molecule sequencing.

This invention relates to counting the absolute number of entities of a species of interest. The invention is based upon a realisation that, while directly counting an absolute number of entities is a challenging and in some cases virtually impossible prospect, it can be relatively straightforward to determine how many different types of entity are present in a sample. For example, as discussed above, molecules having different sequences in a sample can be detected by sequencing, and techniques such as amplification, subtractive hybridisation and/or normalisation of a sample can be used to process the sample before sequencing so that the presence of each different molecule can be reliably and confidently determined.

The present invention relates to counting, measuring or determining the absolute number of entities of a species of interest in a sample by ensuring that the entities of that species of interest differ from each other, and determining the absolute number of the different entities. By ensuring that the entities of the species of interest are detectably different, this allows the absolute number of entities of the species of interest to be determined based on the number of different entities of that species.

Entities of a species of interest can be modified to render them different, e.g. by labelling or other modification as described in more detail below. This facilitates determination of the absolute number of those entities of the species of interest, since one can determine how many differently modified entities of the species are present, and from this information the number of original entities of the species of interest in the sample can be derived.

Alternatively, where a population of entities of a species of interest comprises some entities of the species of interest which are the same and some which differ from one another (for

example, where the species of interest is nucleic acid, and a population comprises examples of the same and different nucleotide sequences), it is possible to ensure that the species of interest differ from one another by taking a sample from that population, where the sample size is sufficiently small that the entities of the species of interest in the sample differ from one another (e.g. where each nucleic acid molecule in the sample has a different nucleotide sequence). The absolute number of entities of the species of interest in the sample can then be determined by determining the number of different entities of the species of interest in the sample. The step of sampling to ensure that the entities of the species of interest differ from one another represents a "bottlenecking" approach, to restrict the number of entities of the species of interest which are subsequently amplified and is subsequently performed.

Use of bottlenecking is of particular value in the context of a method in which the entities of the species of interest in the sample are fragments of larger entities in a starting population, where it is desired to know the relative proportions of the larger entities. By amplifying a limited sample of fragments in which each fragment is different, the absolute number of those different fragments can be determined. Following detection of each different fragment, the fragment can be assigned to a larger entity, i.e. the method can comprise a step of determining the larger entity from which each fragment is derived. This provides information on the relative number of the different larger entities in the starting population. It is useful in a method comprising karyotyping, to determine relative number of different chromosomes.

Alternatively, or additionally, one can ensure that the species of interest are different from one another by providing or generating a plurality of different entities, e.g. randomly different molecular structures. These can, for example, be employed in situations such as tracing or tracking of substances or fluid flow. Thus, where the entities of the species of interest are mixed with a substance to be monitored, samples may be taken after downstream processing, and the amount (e.g. number of entities, or concentration) of the substance present in the sample may be accurately estimated based on the absolute number of entities of the species of interest in the sample.

The invention provides a method of determining the absolute number of entities of a species of interest in a sample, comprising:

ensuring that the entities of the species of interest in the sample differ from each other, and

determining the absolute number of the different entities of the species of interest.

The method may comprise:

(i) ensuring that the entities of that species of interest in the sample differ from each other,

(ii) amplifying the different entities of the species of interest in the sample, and

(iii) determining the number of different entities of the species of interest after amplification, or determining the number of different amplicons,

thereby determining the number of entities of the species of interest in the sample.

Number or concentration can be determined, these being interconvertible based on the volume.

Amplification is performed on the entities in the sample. The sample contains the entities of the species of interest to be counted, which in some cases may be present in modified form in the sample, e.g. as labelled entities or conjugates.

The amplification of the different entities of the species of interest in the sample provides a library of amplicons, each amplicon being a copy of an entity of the species of interest in the sample. The number of different amplicons is representative of the number of entities of the species of interest in the sample before amplification. Thus, the number or concentration of entities of the species of interest in the sample can be determined based on the number or concentration of different amplicons.

The invention provides a way to preserve or store information about the absolute number of entities of a species of interest in a sample (e.g. the number of instances of a particular nucleic acid sequence, or the total number of nucleic acid molecules), so that this information is not lost during processing of the sample by methods such as amplification, subtractive hybridisation and normalisation. By ensuring that the entities of the species of interest in a sample are different from each other before such processing steps are carried out, the absolute number of entities of the species of interest in the sample remains equal to the number of different entities of the species of interest, even after amplification. Processing of a sample (e.g. by amplification, normalisation and/or subtractive hybridisation) may change the relative proportions of entities of the species of interest and of other entities in the sample. But, provided that all entities of the species of interest in the sample were different before amplification, the number of different entities of the species of interest after amplification is representative of the absolute number of entities of the species of interest in the sample before amplification. Thus, a step of ensuring that the entities of the species of interest are different from one another in a sample bestows a form of "molecular memory" on the sample, so that information about the absolute number of entities of a species of interest in that sample is stored, and is not destroyed by subsequent changes in the absolute and relative proportions of species in the sample.

In a first aspect, the invention provides a method for determining the number or concentration of entities of a species of interest in a sample, the method comprising modifying each entity of the species of interest to provide a plurality of differently-modified entities, and determining the number or concentration of entities of the species of interest in the sample, based on the number of differently-modified entities.

The number of differently modified entities is representative of the absolute number of entities of the species of interest which were present in the original sample. Changing the entities of the species of interest into differently-modified entities means that techniques for determining the number or concentration of different species can be used to determine how many (or what concentration of) entities of the species of interest were originally present in the sample. As discussed already above, such techniques are generally easier and faster than attempting to directly determine the absolute number of molecules such as nucleic acids. The invention thus provides an elegant solution to the problem of determining absolute molecular number, and facilitates counting the number of entities of a species of interest in a sample, by enabling the number of differences to be determined instead of directly counting the number of the species of interest. The invention therefore has particular use for counting entities in situations where it would be more difficult or time consuming to count the entities directly, compared with determining how many differently modified entities are present. Furthermore, the methods of the invention may be applied in parallel, where they provide a powerful way to determine the absolute number of many millions of individual species of interest in parallel, using multiplex techniques, and find use in a variety of different applications, as will be explained in more detail below.

The entities may be any items to be counted, provided that they can be differently modified in order to apply the techniques of the invention. In embodiments of the invention which do not require entities to be modified, e.g. where bottlenecking is used to ensure that entities of a species of interest differ from each other in a sample, then it is not necessary that the entities are differently modified, although it may be preferable to do so, and to use modifying and bottlenecking approaches in combination. Entities which are especially suitable are those to which a label can be attached with relatively stable stoichiometry and then the labelled entity purified from free label. Typically, the entities are molecules, e.g. biomolecules such as nucleic acids, proteins, polysaccharides or lipids. The entities may be molecular complexes comprising two or more molecules, which may be two or more biomolecules belonging to the same class or to different classes of biomolecule (classes of biomolecules in this context include nucleic acids, proteins, polysaccharides and lipids). Alternatively, the entities of a species of interest may be comprised in a molecule, rather than being separate molecules. For example, the species of interest may be a specific nucleotide sequence comprised in a chromosome, and the number of

copies or repeats of that sequence in the chromosome may be determined by the method of the present invention.

In the following discussion the term "molecule" is used, but as set out above, it is intended that the invention is not necessarily limited to the determination of the number or concentration of molecules.

The molecules of the species of interest may be modified by labelling, that is, by providing the molecules of the species of interest with variant labels to provide a plurality of variant molecules, or conjugates.

Accordingly, a first development of the first aspect of the invention provides a method for determining the number or concentration of molecules of a species of interest in a sample, the method comprising labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates, each conjugate comprising a label attached to a molecule of the species of interest, and determining the number or concentration of molecules of the species of interest in the sample, based on the number of different conjugates.

Further preferred or optional features of the invention will now be set out. These may be applied singly or in any combination with any aspect or development of the invention described herein, unless the context demands otherwise.

The label is a marker, tag, adapter, part, sequence or structure that serves to distinguish a molecule of the species of interest from another molecule of the species of interest.

The label may be a chemical group or moiety, which may be a polymer, e.g. a polymer comprising a sequence of different units. The label is preferably a nucleic acid (i.e. a polynucleotide), which may be DNA or RNA, or may be a modified nucleic acid such as PNA, LNA or L-DNA. In some embodiments of the present invention, modified nucleic acids may be preferable for use as labels due to their relatively high chemical stability. A nucleic acid label may be a double-stranded or single-stranded nucleic acid

A nucleic acid label may be at least 2 nucleotides in length. Preferably, a nucleic acid label is at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,16 ,17 ,18 ,19 or 20 nucleotides in length, and may optionally be up to 20, 25, 50 or up to 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000 nucleotides in length. A nucleic acid may be 2-100, 2-50, 2-25, 2-20, 4-20, 4-10 or 4-8 nucleotides in length. Preferably a nucleic acid is 4-20 nucleotides in length.

Preferably, the label is a polymer having a random monomer sequence. A random polymer sequence is a sequence comprising more than one type of monomer, in which the sequence of types of monomer is not predetermined by controlled steps during synthesis of the polymer, but depends only on random collisions between reactant monomers and/or between reactant monomers and the nascent polymer. Accordingly, the method of the present invention may comprise a step of generating a group of polymers (such as nucleic acids, or polynucleotides) having random sequences of nucleotides. The generation of a random polymer may comprise the use of N random monomers of B different types, which results in B^N different possible polymers. The generation of a random polynucleotide may comprise the use of N random nucleotides comprising 4 different bases, which results in 4^N different polynucleotides. For example, for a label which is a random polynucleotide synthesised from nucleotide monomers comprising one of four different bases, which label is ten nucleotides in length, there are $4^{10} = 1{,}048{,}576$ different possible sequences. If such a label is twenty nucleotides in length, there are $4^{20} \approx 10^{12}$ (i.e. approximately a trillion) different possible sequences i.e. 10^{12} different possible labels.

Preferably, the group of different labels comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 500, 1000, 2000, 5000, 10,000, 50,000, 100,000, 500,000, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{15} or 10^{18} different labels. Identical labels may be present in the group. There is a statistical probability that such identical labels will be generated, if the labels are generated randomly e.g. comprising random monomer units as discussed above. Accordingly, the group of different labels may comprise two or more labels that are identical to each other.

The methods of the present invention may comprise labelling each molecule of a species of interest with a label, wherein each label is selected from a group of different labels. It may be the case that each label that is selected from the group of different labels may be different from every other label that is selected from the group of different labels. If the group of different labels comprises two or more identical labels, then it may be the case that a label selected from the group of different labels is the same as another label selected from the group of different labels.

The invention provides a method of determining the number or concentration of entities of a species of interest in a sample, the method comprising:

providing the sample from a population of entities of the species of interest, wherein the entities of the species of interest in the sample differ from each other,

amplifying the different entities of the species of interest in the sample to provide a library of amplicons, then

determining the number or concentration of different amplicons, and thereby determining the number or concentration of entities of the species of interest in the sample.

The population may comprise identical and different entities of the species of interest. The size of sample taken may be adjusted such that the entities of the species of interest in the sample differ from each other.

The sample may be provided by physically separating the sample from the population, e.g. as an aliquot. Alternatively, the sample may be provided in same reaction mixture, for example by ensuring that amplification can only be performed on some of the entities of the species of interest in the population, e.g. by using a limited number of adaptors or label molecules. The provision of a sample represents a bottlenecking step.

The method may comprise the bottlenecking step in combination with a modifying step. For example, the method may comprise:

modifying entities of the species of interest in a population to provide a plurality of differently-modified entities,

providing the sample from the population, wherein the entities of the species of interest in the sample differ from each other,

amplifying the different entities of the species of interest in the sample to provide a library of amplicons, then

determining the number or concentration of different amplicons, and thereby determining the number or concentration of entities of the species of interest in the sample.

The modifying step provides a plurality of differently modified-entities, so that at least some entities of differently modified, e.g. all or almost all entities may be differently modified. Optionally, there may be some entities in the population which are identically modified, or which are not modified, so that the population comprises some identical and some different entities of the species of interest. For example, this may be the case where the modifying comprises labelling molecules in the population with fewer labels than the number of molecules. In such a situation, the step of providing a sample can reduce the number of identical entities of the species of interest, e.g. the number of identically modified entities of the species of interest, so that the entities of the species of interest in the sample differ from each other. Alternatively, the step of modifying entities of the species of interest can be performed such that all of the entities of the species of interest differ from each other, in which the entities of the species of interest will also differ from each other in the sample that is provided from the population.

As discussed elsewhere herein, it is not essential that absolutely all entities of the species of interest are rendered mutually distinguishable, since the method can still be applied when almost all entities of the species of interest differ. For example, at least about 90%, 95%, 98%, 99 %, 99.9 % or 99.99 % of the entities of the species of interest may differ from each other.

In accordance with the second aspect of the present invention, there is provided a method of tracking a population of entities of a species of interest that comprises providing a plurality of entities that are different from each other, as set out in more detail below. The plurality of different molecules may be polymers e.g. nucleic acids, which may be synthesised such that they are each different from each other, for example such that they have different sequences from each other (e.g. 20bp random sequence). In this context the different sequence structures of the nucleic acids are effectively labels, even though no step of attaching a label is performed.

Methods in accordance with the second aspect of the invention may comprise generating a first plurality of first molecules and generating a second plurality of second molecules, as discussed below. The first and second molecules may be nucleic acids, which may be synthesised so that they have molecule-identifying sequence (e.g. 20bp random sequence, or label) and a plurality-identifying sequence (e.g. 20bp predetermined sequence) which is different between different pluralities of molecules, but may be the same within a particular plurality of molecules.

Different labels may differ from each other in terms of length and/or sequence. The labels may be intentionally made different from each other by more than one monomer (e.g. more than one bases) to improve unambiguous detection of the labels and to decrease the effect of amplification and sequencing errors. The advantages of using labels of different lengths, e.g. to allow detection of the different labels by mass spectrometry, is further discussed below.

The species of interest is the defined physical entity to be counted, e.g. a specific or defined molecular structure. The species of interest may be a class of biomolecule, such as nucleic acid, or protein. The species of interest may be a specific protein (i.e. polypeptide). Preferably the species of interest is a nucleic acid comprising or consisting of a specific nucleotide sequence. A nucleic acid species of interest may be DNA, for example cDNA, or may be RNA, for example mRNA (messenger RNA) or miRNA (microRNA). A nucleic acid species of interest may be single-stranded or double-stranded.

The species of interest may be a distinct type, a single type, or an individual kind, of molecule or other entity in the sample. The species of interest may be a nucleic acid comprising or consisting of a nucleotide sequence that is distinct in the sample (e.g. a specific nucleotide sequence that is present only in a sub-set of the nucleic acids in the sample), and the method of

the invention is used to determine the number of nucleic acid molecules in the sample comprising or consisting of that distinct nucleotide sequence. A molecule of a species of interest which is a nucleic acid comprising a specific nucleotide sequence may be a chromosome, or chromosome fragment, which comprises a specific nucleotide sequence of a gene of interest and/or which comprises a specific nucleotide sequence of a mutation of interest. The method of the invention allows the number or concentration of chromosomes (which may differ from each other) that comprise that specific nucleotide sequence to be determined. A species of interest which is a nucleic acid consisting of a specific nucleotide sequence may be a messenger RNA. The method of the invention allows the number or concentration of mRNAs that consist of a specific nucleotide sequence to be determined.

The species of interest may be a defined sequence or structure occurring within a larger sequence or structure, for example a defined sequence of monomers comprised within a polymer. The species of interest may be a specific nucleotide sequence, or region, comprised in a genomic sequence or chromosome. Different sequences in the same chromosome can be individually and separately counted, as explained in more detail below. This is relevant for example for copy number variation and deletion analysis where number of sequences within different parts of the same molecular structure (chromosome) are analyzed. Thus the present invention encompasses determining the number of copies of a sequence of interest in a sample.

A chromosome may have regions that are duplicated (2 or more copies of the same sequence), deleted (0 copies) or translocated (0 copies of original sequence, 1 or more copies of novel sequence). Such deletions, duplications or "amplifications" are very common in hereditary diseases and in cancer cells (e.g. *http://www.ncbi.nlm.nih.gov/cancerchromosomes*).
Parts of chromosomes can be detected in many ways, for example using PCR with specific primers to amplify a region, or cutting the chromosome and analyzing the fragments. If the PCR products or fragments are labelled in accordance with the present invention, then the number of labels indicates the number of copies of a particular sequence in the chromosome.

Thus, where the species of interest is a molecule comprising or consisting of a defined molecular structure, this may be a single exact molecular structure, in which case the method of the invention comprises determining the number or concentration of exact copies, i.e. identical versions, of this species of interest.

Where the species of interest is a molecule comprising or consisting of a molecular structure permitting possible variation, e.g. isomerisation, different chirality, and/or substitution of molecular groups, then the method of the invention comprises determining the number or concentration of the molecular structure which optionally encompasses a family of closely-

related structures. For example, the method can be used to count DNA molecules which have non-identical but closely related sequences by inclusion of bases such as inosine which hybridize with multiple nucleotides.

In some embodiments, a species of interest may be a nucleic acid, or nucleotide sequence, comprising or consisting of any member of a family of similar nucleotide sequences, and the method of the invention is used to determine the number of nucleic acid molecules in the sample comprising or consisting of a member of that family of similar nucleotide sequences. The family of similar nucleotide sequences may be closely-related non-identical sequences. Closely-related non-identical sequences may be defined as sequences that share at least 70%, 80%, 90%, 95%, 98%, or 99% sequence identity.

According to the present invention, a molecule of a species of interest may be labelled to provide a conjugate. In the context of the present invention, the term conjugate refers to any compound comprising a molecule of a species of interest associated with a label. Preferably, a conjugate comprises a molecule of interest moiety covalently linked to a label moiety. However, a label and molecule of the species of interest in a conjugate can be associated together by any means. For example, by enzymatic ligation, by template-directed polymerisation, by topoisomerisation, using an integrase or recombinase, or by chemical ligation.

The label can be associated with a molecule of the species of interest to provide a conjugate by any suitable means. For embodiments in which the molecule of the species of interest is a nucleic acid, the label may be covalently attached to the 5' or the 3' end of the nucleic acid, or inserted at any position between the 5' or 3' end.

The species of interest may belong to a genus of molecules, such that the molecules of the species of interest are molecules of a genus of molecules. Labelling each molecule of the species of interest in to provide a plurality of conjugates, in accordance with the present invention, may comprise labelling each molecule of the genus of molecules in the sample, to provide a population of labelled molecules comprising the plurality of conjugates.

Preferably the species of interest is a nucleic acid having a specific sequence, and the genus of molecules is a genus of nucleic acids having various sequences. The method of the invention may comprise labelling each nucleic acid molecule in a sample, to provide a plurality of labelled nucleic acids, which plurality of labelled nucleic acids comprises a plurality of conjugates, which conjugates each comprise a nucleic acid having the specific sequence and a label.

The method of the present invention may include the additional step of determining the total number of molecules in the sample belonging to the genus of molecules. The method of the

invention may comprise labelling each molecule of the genus of molecules in the sample, and the step of estimating the total number of molecules in the sample belonging to the genus of molecules, based on the sampling distribution of observed unique (i.e. single copy) and non-unique (i.e. multiple copy, e.g. duplicate, triplicate etc) labels in the population of labelled molecules.

The methods of the present invention preferably comprise amplifying the different, or differently-modified, molecules to provide a library of amplicons. Preferably, the methods of the invention comprise the step of amplifying the part of the conjugate comprising the label to provide a library of amplicons, wherein the number of molecules of the species of interest in the sample is determined based on the number of different amplicons.

Amplifying part of the conjugate comprising the label encompasses amplifying only the label (or a part thereof), or amplifying the label (or a part thereof) and the conjugate (or a part thereof). Amplifying part of the conjugate comprising the label may comprise amplifying the entire conjugate. Hence a library of amplicons may comprise amplified different molecules, amplified differently-modified molecules, amplified labels, amplified parts of labels, amplified parts of conjugates, or amplified conjugates.

The term amplification refers to processes that replicate a molecule, and/or increase the number of copies of a molecule. Amplification may comprise the polymerase chain reaction.

A nucleic acid label may include forward and/or reverse primer sites to facilitate amplification by PCR. A nucleic acid label may comprise a forward and reverse primer site, wherein the forward and reverse primer sites flank a random nucleotide sequence. The nucleic acid label may comprise a single forward or reverse primer site and a random nucleotide sequence, for example when the species of interest is also nucleic acid, in which case the complementary reverse or forward primer site may be formed by a nucleic acid sequence in the species of interest.

The methods of the present invention may comprise normalising the plurality of differently-modified molecules, plurality of conjugates, or the library of amplicons. Preferably, the methods of the invention comprise normalising a population of differently-modified molecules or conjugates that has been amplified.

The term normalisation refers to processes that normalise the relative abundance of molecules of different species in a sample, that is, processes that increase the relative abundance of rare species and/or decrease the relative abundance or common, or dominant, species in a sample.

Normalisation may comprise subtractive hybridisation. Normalisation methods are known in the art (see e.g. Soares et al., 1994; Coche et al., 1994).

The step of determining the number or concentration of molecules of the species of interest in the sample may be based on the number of different amplicons present in the library, which library may have been normalised. Preferably, the method of the invention comprises a step of obtaining an aliquot from the library of amplicons, which library has preferably been normalised, and the step of determining the number or concentration of molecules of the species of interest in the sample is based on the number of different amplicons present in the aliquot.

Preferably, following the step of labelling the molecules of the species of interest, the method of the present invention comprises a step of purifying, or separating, conjugates from free label, before the conjugates are further processed, or the number or concentration of molecules of interest in the sample is determined.

It is desirable to ensure that any free label which remains after the labelling reaction is not amplified, since this would adversely affect the counting accuracy. In methods of the invention, it is preferable that labels are not amplified when they are in free form and not in the form of a conjugate. This can be achieved for example by using labels that are not amplifiable in free form, or by separating conjugates from free label that is present after labelling. For example, free label can be chemically removed (e.g. degraded) and/or physically separated from the conjugates. Where the free label is amplifiable, e.g. where it is nucleic acid, a separation step can be performed before amplification. For example, RNA labels can be used, and converted to DNA using reverse transcriptase after the labelling reaction, while free labels are degraded using RNase.

The method of the invention stores information about the original number of molecules of a species of interest in a sample. This information about the original number of molecules of a species of interest is preserved during subsequent amplification and/or normalization steps. Thus, the method of the invention allows accurate counting of the original number of molecules of a species of interest in a sample even after extensive processing steps (e.g. amplification and/or normalisation) have been performed, for example to improve detection of rare species of interest (or for any other reason).

The method of the invention overcomes the disadvantages associated with methods of the prior art that involve amplifying, and/or normalising, the molecules in the sample (e.g. amplifying mRNA molecules in a sample by PCR). In such prior art methods, information about the original number of molecules of a species of interest in the sample is (partially or completely) destroyed. This is a disadvantage because to recover information about the presence of molecules of a

rare species of interest in a sample, it is desirable to amplify (increase the copy number of) that species, but in doing so information about the original number of molecules of that species is irretrievably lost. One reason for this loss of information is that the degree or rate of amplification depends on the structure of the molecules being amplified. For example, the number of molecules of nucleic acid amplified by a specific number of rounds of PCR depends not only on the original number of molecules of nucleic acid, but also on the sequence and the length of the nucleic acid (e.g. shorter molecules of nucleic acid are amplified more efficiently that longer molecules of nucleic acid). Similarly, efficiency of normalisation of populations of molecules nucleic acids depends on their sequence and length. Prior art methods for determining the number of molecules of a species of interest in a sample that involve amplification and/or normalisation are therefore unreliable and/or inaccurate.

By contrast, the method of the invention stores information about the original number of molecules of a species of interest in the sample by ensuring that the molecules differ from one another, e.g. modifying those molecules to produce a plurality of differently-modified molecules. The original number of molecules of the species of interest in the sample can be determined based on the number of different differently-modified molecules, or the number of different molecules. The population of differently-modified molecules can be amplified and/or normalised such that differences between them are preserved, and therefore information about the original number of molecules in the sample is preserved during subsequent rounds of processing of the sample, such that it is possible to accurately determine the number of molecules of interest in a sample, even when the population of molecules from the sample has been amplified and/or normalised.

The method of the invention overcomes the disadvantages associated with methods of the prior art that involve detecting every individual molecule in a sample (e.g. single molecule sequencing methods). The total number of molecules in a sample may be extremely large, in which case detecting every molecule in the sample is laborious. If the species of interest is relatively rare, then it is necessary to detect a very large number of molecules in order to count a relatively small number of molecules of a species of interest, which is inefficient.

By contrast, the method of the invention stores information about the original number of molecules of a species of interest in the sample, e.g. by modifying those molecules to produce a plurality of differently-modified molecules. Therefore it is possible to process the sample by amplification and normalisation to determine the number of molecules of the species of interest in the sample without detecting the large number of molecules present in the unprocessed sample.

The method of the present invention preserves information about the number of molecules of a species of interest in a sample by modifying molecules of the species of interest so that they differ from one another. This means that in order to determine the number of molecules of the species of interest, instead of having to enumerate every one of a group of identical molecules, one can count the number of differently modified molecules. The method of the present invention depends on the distribution of different modifications within the molecules of the species of interest, as explained in more detail below, so even if each molecule is expected to be differently modified (for example, when each molecule is labelled with a label selected from a group of different labels), it is possible that some molecules of the species of interest may be modified in the same way just by chance (e.g. two or more molecules of the species of interest are labelled with identical labels selected from the group of different labels). However, provided that the fraction of molecules modified in the same way is small in relation to the total number of molecules of the species of interest, the method of the invention will provide an accurate determination of the number of molecules of the species of interest. The probability that two or more molecules of the same species of interest will be modified in the same way is lower for species of interest that are present in the sample in low number, and lower for larger groups of modification (e.g. groups of different labels having a larger number of different labels). Therefore, if the method of the present invention is used to determine the number or concentration of molecules of a species of interest that is rare in the sample and the group of modifications used to modify the molecules is sufficiently large, then the chances of more than one molecule of the species if interest being modified in the same way are non-significant or negligible. In other words, the methods of the present invention do not require that each molecule of the species of interest is modified such that it differs from every other molecule of that species of interest.

For maximum accuracy, and to avoid the problem of under counting which may occur if two or more molecules of the same species of interest in a sample are identical, e.g. if they are are by chance labelled in the same way (i.e. labelled with identical labels) to provide two or more identical conjugates, it is preferable to optimise the probability that every entity of the species of interest in the sample to be counted is rendered unique, i.e. that it is distinguishable from every other entity of the species of interest in the sample. This is particularly relevant when amplification and/or normalisation steps are to be performed on the sample. If two or more identical entities of the species of interest were amplified, the resulting identical amplicons would be counted as representative of only one entity.

So, preferably a sample is provided in which all, or almost all, of the entities of the species of interest are mutually distinguishable, i.e. unique. For example, at least about 90%, 95%, 98%, 99 %, 99.9 % or 99.99 % of the entities of the species of interest may be unique.

The invention provides ways to maximise the proportion of entities that are unique, and thus to reduce the probability that two or more identical molecules are amplified, and so improve the counting accuracy.

One approach is to randomly modify each entity of the species of interest in the sample, including enough random modifications to render each entity unique. For example, where the method comprises labelling molecules of a species of interest, it is desirable to provide enough random labels so that each entity is differently labelled. The number of labels required is a function of the number, or expected number, of originally identical molecules in the sample. So, in a sample containing thousands of genomes, thousands of different labels would be needed. If the number of identical molecules of the species of interest in the sample is *n* then the minimum number of different labels needed to provide a population of conjugates in which every conjugate is unique is *n-1* (since one molecule of the species of interest may be distinct from the other molecules by having no label).

Another approach is to modify each entity of the species of interest in the sample non-randomly. This can be accomplished by synthesizing a defined set of labels, and/or by arraying, ordering or binding together multiple labels physically or chemically prior to or during the labelling reaction, for example by using polymers (including DNA), macromolecules, particles, beads or microarrays.

For simplicity, we describe this in the context of modifying molecules by labelling to produce label-molecule conjugates which represent differently modified molecules. However, the principle applies to types of modification other than labelling, and to other types of modified molecules, in accordance with the disclosure herein.

Molecules of a species of interest may be labelled with a label selected from a group of different labels to provide a plurality of conjugates, wherein the number of different labels is larger than the number of molecules of the species of interest which are labelled. The larger the number of different labels, the lower the probability that two or more molecules of the species of interest will be labelled with the same label. The number of different labels may be about, or at least about 1, 1.2, 1.5, 2, 3, 4, 5, 10, 20, 50, 100, 500, 1000, 10,000, or 100,000, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{15} or 10^{18} times larger than the number of molecules of the species of interest. The number of different labels may be about 2 – 100,000, 2 – 1,000, 2 – 100, 2 – 10, 10 – 10^6, 10 – 10,000, 10 –1,000, 1,000 – 10^9, 1,000 – 10^6, or 1,000 – 100,000 times larger than the number of molecules of the species of interest. Preferably, the number of different labels is

sufficiently larger than the number of molecules of the species of interest to ensure that all, or almost all, of the conjugates is unique.

Another approach is to use fewer different labels than molecules of the species of interest to be labelled, and then subsequently take one or more sufficiently small aliquot(s) in which the vast majority of molecules are different. For example, one may provide a sample by taking one or more aliquots from a population, e.g. from serial dilutions of a population of labelled molecules.

For example, labelling 10,000 molecules with 500 different labels, and then taking 1% of this approximately 100 labelled molecules (conjugates). The vast majority of these would be unique and could be counted accurately using the method of the invention. For example, by amplifying the 100 conjugates and determining the number of molecules of the species of interest in the aliquot based on the number of different amplicons. The number of different molecules in the population from which the sample was taken can be estimated based on e.g. the relative volumes of the population and sample. Counting the molecules using the method of the invention based on the number of differently labelled molecules in the whole population would be less accurate, because there are many more molecules of the species of interest than different labels used to label them, and therefore the whole population includes an undesirably high proportion of identically modified molecules of the species of interest (identical conjugates).

A method in which the number of different labels used is lower than the number of molecules of the species of interest may be advantageous because it is often technically simpler to modify all of the molecules in a population and then dilute the population, take a sample of the diluted population, and determine the number of differently modified molecules in the sample.

The population of modified entities may be statistically more likely to contain entities which are by chance modified in the same way, compared with the smaller sample. The sample can be provided such that each entity of the species of interest differs from each other (at least, as explained elsewhere herein, it can be arranged so that this is statistically likely to be the case, even if it cannot always be avoided that by chance that some identically modified molecules may yet be present even in a very small sample).

Another advantage of reducing the size of the sample is that the library of amplicons generated from that sample contains fewer different types of amplicon. It can be faster and more economical to count the different types of amplicon in a library obtained from a smaller sample, and this increases the confidence and reliability of the counting. For example it can reduce the depth of sequencing required for counting different types of nucleic acid molecules.

Accordingly, the sample comprising the entities of the species of interest to be counted may be provided from a larger population of entities of the species of interest. This represents a "bottlenecking" step. In some cases, this avoids the need to modify the entities of the species of interest, because the step of providing the sample from the population can itself ensure that the entities of the species of interest in the sample differ from one another. This is the case where, for example, the species of interest is a class of molecules encompassing different molecular structures, rather than one exact molecular structure. A sample can be provided from the population, so that the entities of the species of interest in the sample differ from each other, as discussed above.

A step of providing a sample need not require physically taking an aliquot from a population, although this is one possibility. Where amplification is to be performed, the sample can be provided by restricting in any other way (e.g. physically or chemically) the proportion of entities of the species of interest on which the amplification step is performed. For example, a bottleneck could be introduced by using an intentionally limited amount of adaptors for an amplification reaction, or by using suboptimal reaction conditions when preparing the sample, or by excising a very small amount of DNA from a gel. A step of providing the sample comprising the entities of the species of interest to be counted may thus reduce the number of entities of the species of interest in the sample compared with the number in the population.

A bottlenecking step reduces the number of molecules of a species of interest to be counted. For example, a bottlenecking step may provide a sample containing fewer molecules of the species of interest than the population. Alternatively or additionally, a bottlenecking step may provide a sample in which relatively few molecules of the species of interest are capable of being amplified, or in which relatively few molecules of the species of interest are labelled with a label capable of being amplified. A bottlenecking step may reduce the number of molecules by a reduction factor of about, or at least about, 2, 3, 4, 5, 10, 20, 50, 100, 500, 1000, 10,000, 10,0000, or 100,000, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{15} or 10^{18}. A bottlenecking step may comprise taking an aliquot, portion, sample, sub-sample or sub-plurality from a volume of solution, which solution comprises a population, plurality of different molecules, or plurality of conjugates. The aliquot, portion, sample, sub-sample or sub-plurality may have a volume that is about, or at least about, 2, 3, 4, 5, 10, 20, 50, 100, 500, 1000, 10,000, 10,0000, or 100,000, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{15} or 10^{18} times smaller than the volume of the solution.

A bottlenecking step may be advantageous because, for example, it may reduce the chance that two or more molecules of a species of interest are identically modified and then amplified, which is undesirable because it reduces the accuracy of counting of the molecules of the species of interest. When the method of the invention includes a bottlenecking step, the

number of different amplicons indicates number of molecules of the species of interest present in the sample after the bottlenecking step. If the reduction factor of the bottlenecking step is known then the number of molecules of the species of interest present in the sample before the bottlenecking step (e.g. in the population) may be determined.

By definition, the number of molecules of a species of interest in a sample is unknown before the method of the invention is used to count them. However, the number of different labels in the group of different labels may be adjusted based on, for example, the expected number of molecules of the species of interest in the sample, or the approximate total number of molecules in the sample (e.g. estimated from spectroscopic or biochemical analysis of the total concentration of the genus biomolecule to which the species of interest belongs, such as nucleic acid). For instance, assuming that the sample does not consist entirely of identical molecules of the species of interest, if the number of different labels is greater than the total number of molecules in the sample, then the number of different labels will be greater than the number of identical molecules of the species of interest.

Methods comprising modifying entities of a species of interest, and methods comprising providing a sample, may be combined. For example, a method may comprise modifying the entities of the species of interest in a population so that all, or almost all, are distinguishable from each other (e.g. by using a sufficiently high number of labels), and then including a bottlenecking step, such as taking an aliquot or providing a sample in another way, so that subsequent steps (e.g. amplification) are performed on the sample of the population. This provides the advantage that the number of different amplicons can be determined more quickly and efficiently, and also helps to ensure that all entities of the species of interest in the sample are unique.

Where the sample is obtained from a population, a method may further comprise determining the number of entities of the species of interest in the population.

Where the entities of the species of interest in the sample are fragments of larger entities, the method may further comprise determining the relative number of the larger entities. For example, chromosomes may be fragmented to produce a population of fragments, and a sample can be provided from that population, ensuring that the fragments in the sample differ from each other. The number of fragments in the sample is then determined, and karyotyping can be used to determine the chromosome from which each different fragment was derived. The relative number of different chromosomes represented in the sample, and thus in the population, can thereby be determined.

Accordingly, a method of the invention may comprise determining the relative number of two or more different chromosomes, comprising

providing a population of fragments of the chromosomes,

providing a sample from the population of chromosome fragments, wherein the chromosome fragments in the sample differ from each other,

amplifying the different chromosome fragments in the sample to provide a library of amplicons,

determining the number of different amplicons, thereby determining the number of different chromosome fragments in the sample,

identifying the chromosome from which each fragment in the sample was derived, thereby determining the relative number of the different chromosomes represented in the sample.

Methods of determining the number of different amplicons are described herein. For chromosome fragments, sequencing can be used to determine the number of different amplicons.

The chromosomes may, for example, be provided from an individual. The method may comprise determining the relative number of two or more different chromosomes in the genotype of an individual.

The method may comprise digital karyotyping, which is also known as virtual karyotyping. Accordingly, the sample preferably comprises fewer than n genomes, where n is the length of chromosome fragments. This maximises the probability that each fragment, and therefore each sequencing read of the different amplicons, has a distinct starting position.

The method may comprise paired-end sequencing. Accordingly, the sample preferably comprises fewer than kn genomes (where n is the length of chromosome fragments and k is the expected distribution of mate positions). This maximises the probability that each read-pair has a distinct combination of first and second mate positions.

Chromosomes in a specimen obtained from an individual, e.g. a human, are typically obtained in fragmented form. It therefore may not be necessary to include deliberate processing steps to cause fragmentation. However, if desired, a step of fragmenting the chromosomes may comprise physical fragmentation (e.g. sonication) or chemical fragmentation (e.g. restriction enzyme digestion). The chromosome fragments may have an average length of about 50 bp, 100 bp, 150 bp, 200 bp, 500 bp or 1000 bp, and a fragmentation step may be adapted to provide fragments of this length.

An embodiment of the present invention is a virtual karyotyping method involving determining the relative number of two different chromosomes in the genome of an individual, e.g. a human. This method may involve determining the relative number of two different chromosomes in a specimen, or biopsy, obtained from that individual. Such a specimen may have been obtained from the individual directly, or indirectly (e.g. where the individual is a foetus and the specimen is obtained from the mother of the foetus).

The specimen may contain chromosomes representing many complete genomes of that individual, and therefore the specimen may contain many identical versions of the two different chromosomes of interest. The method of the invention comprises fragmenting the chromosomes to provide a plurality of chromosome fragments. Thus, from a specimen containing many identical versions of two different chromosomes, a population is provided containing relatively few identical versions of many different chromosome fragments. By performing a bottlenecking step, which provides a sample containing a relatively small number of fragments, a sample is provided in which the chromosome fragments differ from each other. For example, a sample of genomic DNA fragmented to 200 bp may initially contain billions of molecules (corresponding to thousands of genomes), many of which are accidentally identical. By taking a small aliquot, say corresponding to a few million molecules, one can be certain that nearly all molecules are distinct. The introduction of a bottleneck helps to ensure uniqueness. The fragments in the sample are then amplified, and may optionally be normalised and/or sequenced, and assigned to one of the two different chromosomes (i.e. the chromosome from which they are derived is identified). Because the bottlenecking step eliminates identical fragments, the number of fragments assigned to each of the two different chromosomes is representative of the relative number of fragments from each of the different chromosomes, which is in turn representative of the relative number of the different chromosomes in the specimen. Therefore the relative number of chromosomes represented in the sample is determined based on the relative number of fragments assigned to each respective chromosome.

This embodiment of the present invention may be described with reference to Figure 4, in which two different chromosomes of interest are represented as a white chromosome and a black chromosome respectively. The original specimen comprises many identical versions of each of the two different chromosomes. Fragmentation provides a population comprising a large number of fragments, a minority of which will be identical. Different fragments of the white chromosome are represented by different numbers, and different fragments of the black chromosome are represented by different letters. The bottlenecking step provides a sample in which each fragment is unique. These fragments may then be amplified and assigned to each

of the two different chromosomes, each different amplicon being representative of a single fragment in the sample. Fragments may be assigned to a chromosome based on their nucleotide sequence, which may be determined by a sequencing step following the amplification step. The relative number of "hits" for each chromosome (fragments assigned to each chromosome) is representative of the relative number of the two different chromosomes in the specimen. This method can be used to diagnose conditions caused by chromosomal abnormalities. In the schematic method shown in Figure 4 the number of "hits" for each chromosome indicates that the relative number of white to black chromosomes is 3 to 2 – this type of result would indicate that the individual from which the specimen is obtained has trisomy for the white chromosome.

In some embodiments the larger entity is a chromosome or a larger stretch of DNA, e.g. part of a chromosome. The method may be used to determine the relative number of two or more stretches of DNA, e.g. two or more different chromosomal regions. The method can be used to diagnose genetic disorders involving chromosomal aberrations such as altered number of one or more chromosomes or duplication of a chromosomal region. For example, the method may be used to determine whether multiple copies of chromosome 21 are present the genome of in an individual, which is useful as a diagnostic test for Down syndrome. This may be used in the context of pre-natal diagnosis, obtaining a sample of DNA from the mother's blood, in which fetal DNA is present. This is explained and exemplified further in Example 4.

Although described here for chromosomes and chromosome fragments, it is to be understood that the method may be applied where the entities of the species of interest are fragments of any larger entity. The entities are fragmented to provide a population of fragments, from which a sample is taken. The population comprises different fragments, however it also comprises some fragments which are the same. The step of providing the sample ensures that the fragments in the sample differ from each other.

As illustrated by the example above, a bottlenecking step can be used to provide a sample in which entities of the species of interest differ from each another, without requiring the entities to be modified to render them different.

However, in many cases it is necessary or useful to include a step of modifying the entities of the species of interest, to ensure that the entities of the species of interest in the sample differ from each other.

Labelling can be used to differentiate molecules of a species of interest, enabling previously identical molecules of a species of interest to be distinguished as different conjugates.

Molecules of a species of interest can also be distinguished according to other factors, such as their length. This reduces the number of different labels which are required to distinguish the entities of the species of interest.

For example, truncation of a nucleic acid molecule can generate a population of molecules of variable length, but within the population there will be multiple truncated molecules which are the same. A sample can be provided in which the truncated molecules are labelled to produce conjugates, which can be distinguished according to their different length or different label. Thus, even if the same label is conjugated to different truncated molecules, the two conjugates are different from each other because of their different lengths or because the label is attached to a different nucleic acid sequence (e.g. the 5' end has a different sequence, owing to the truncation). Amplification of the labelled fragments and sequencing of the amplicons is then used to determine the number of different amplicons. Each different amplicon corresponds to one truncated molecule from the sample, so the number of truncated molecules in the sample can thereby be determined. Further, the number of nucleic molecules which were present in the population before truncation can be determined, since each truncated molecule corresponds to one such nucleic acid molecule.

This method is especially suitable for determining number or concentration of mRNA molecules, which can be progressively truncated from the 5' end to generate different 5' sequences.

An advantage of the methods of the present invention is that they do not require the use of prior art methods to determine the number or concentration of the species of interest, and therefore do not have the disadvantages associated with those prior art methods. Accordingly the present invention provides methods for determining the number or concentration of molecules of a species of interest that do not comprise detecting each individual molecule of the species of interest, and/or do not comprise single molecule sequencing, and/or do not comprise microarray hybridisation, and/or do not comprise sequencing of tags derived from the species of interest.

In accordance with the invention, the number of entities of a species of interest can be determined from the number of different amplicons, following amplification of the sample of different entities of the species of interest.

Determining the number of different amplicons may comprise detecting, e.g. sequencing, the different amplicons. Preferably, a sufficient number of amplicons is detected or sequenced such that each different amplicon is observed at least twice. The more replicates of each amplicon that are observed, the better, e.g. each amplicon may be detected hundreds of times. This is desirable since it increases the confidence of having detected all the different amplicons originating from the different entities of the species of interest in the sample.

Preferably the amplicons are detected or sequenced to a depth such that at least some amplicons are observed more than once. Preferably, most amplicons are observed more than once, and most preferably all amplicons are observed more than once. Preferably, amplicons are observed more than once at least as duplicates. Preferably, amplicons are observed more than once at least as replicates, where the number of replicates is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000. Preferably, at least one amplicon is observed as a replicate. More preferably most, or all, amplicons are observed as replicates. Amplicons may be observed until a threshold number of amplicons are observed as replicates, which threshold number may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000. Amplicons may be observed until a threshold proportion of amplicons are observed as replicates, which threshold proportion may be 1%, 5%, 10%, 20%, 20%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99.5%, 99.9%, or 100%. The higher the proportion of amplicons that are observed at least as duplicates (i.e. at least as 2 replicates), and/or the higher the number of replicates of amplicons that are observed, the more likely it is that the number of molecules of the species of interest determined using the method of the invention is very accurate.

If the number of replicates observed is low, relative to the number of singlicates observed, then the number of molecules of the species of interest may be estimated based on the Poisson distribution i.e. by assuming that replicates are Poisson-distributed, a Poisson model may be fitted to the observed numbers of replicates and singlicates in order to estimate the number of molecules present before the amplification step and thus the number of molecules of the species of interest in the sample. Methods of fitting a Poisson distribution to an observed data set are known (e.g. Stasinopoulos and Rigby). Preferably a zero-truncated Poisson distribution is fitted to the data.

The depth of detecting or sequencing refers to the number of amplicons, relative to the total number of amplicons, that must be detected or sequenced before a sufficient number amplicons are observed in replicate that one can have confidence that the number of molecules of the species of interest determined using the method of the invention is very accurate.

As noted above, a bottlenecking step may reduce the depth of sequencing required. In the absence of a bottleneck step, excessive depth of sequencing may be required to begin observing some amplicons twice. For example, if a sample containing thousands of genomes is labeled with random labels, and then amplified, one must still generate thousand-fold coverage to begin seeing some amplicons twice. Thus a bottleneck may be used to reduce the depth of

sequencing required, and the reduction factor of the bottleneck may be adjusted in view of the number of reads (i.e. detection or sequencing events) that can be economically obtained, and to the desired final accuracy.

Generally, unless context dictates otherwise, in the context of the present invention the terms "duplicate", "triplicate", "replicate" etc. refer to molecules that are copies of each other, in particular amplicons that are the same. These replicates are generated by an amplification (i.e. copying) step in the method of the invention. When the number of different amplicons is determined according to the invention, it is desirable that many replicates are observed of many different amplicons, because then one can have greater confidence that sufficient amplicons have been detected or sequenced in order to determine the number of different amplicons. By contrast the term "identical" is used to refer to molecules that are not copies generated in a step of the method of the invention but are nevertheless the same (e.g. two or more molecules of a particular species of interest; or two or more molecules of a particular species of interest that have been identically modified), that is, molecules in the sample that are the same before any copying/amplification step is performed). As described herein, the amplification of identical molecules is preferably avoided, and therefore the methods of the present invention provide a variety of ways of processing samples to reduce the number of identical molecules.

The method of the invention can be applied in a multiplexed fashion to determine the number or concentration of molecules of two or more different species of interest in a sample. Alternatively or additionally to this multiplexing method, the present invention can be applied in a multiplexed fashion to count the number of molecules of a species of interest present in two or more different samples.

Accordingly, in a second development of the invention there is provided a method for determining the number or concentration of molecules of both a primary species of interest and a secondary species of interest in a sample, the method comprising labelling each molecule of the primary and secondary species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates comprising primary conjugates and secondary conjugates, each primary conjugate comprising a label attached to a molecule of a primary species of interest, and each secondary conjugate comprising a label attached to a molecule of a secondary species of interest; and determining the number or concentration of molecules of the primary species of interest in the sample based on the number of different primary conjugates, and determining the number or concentration of molecules of the secondary species of interest in the sample based on the number of different secondary conjugates.

In some embodiments of the second development of the first aspect of the present invention, the primary species of interest is a nucleic acid having a first specific nucleotide sequence and the secondary species of interest is a nucleic acid having a second specific nucleotide sequence that is distinct from the first specific nucleotide sequence.

In a preferred embodiment of the second development of the first aspect of present invention there is provided a method for determining the number or concentration of molecules of each of a plurality of particular species of interest in a sample, the method comprising labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of particular conjugates, each particular conjugate comprising a label attached to a molecule of a particular species of interest, and determining the number or concentration of molecules of each particular species of interest in the sample, based on the number of different particular conjugates.

Preferably, the plurality of species of interest are a plurality of nucleic acids each having a specific nucleotide sequence.

The plurality of species of interest may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 1,000, 10,000 or 50,000 species of interest. For example, the method of the invention may be used to determine the number or concentration of tens of thousands of different nucleic acids of interest.

It is significant advantage of the present invention that the same labels can be used to label molecules of different species of interest, provided that molecules of the different species of interest can be separated from each other (physically, chemically or by their sequence). For example two different regions of the same chromosome can be labelled with the same set of labels. The number of both regions can then be independently analyzed by first separating the chromosomal regions to different bins by sequence and then analyzing the labels separately in each bin). The key thing is that each molecule in the sample becomes different - even if two species of interest are labelled with the same sequence, it is possible to tell which is which as long as the species of interest themselves have different sequence. Alternatively, the species of interest can be separated chemically before analysis.

The labels can be the same for each of the species of interest, as long as all molecules of interest become different after labelling. Two different species of interest can be separated from each other by sequence derived from the species of interest (or by chemical separation) even if the label sequences used are the same.

For example, as set out below for a DNA-counting application:

LABEL--species of interest
AATA--TAGGAATAGAA
AATC--TAGGAATAGAA
AATA--CAGACAGATAA
TTCC--CAGACAGATAA

Even if both the TAGGAATAGAA and CAGACAGATAA are labelled by "AATA", one can tell them apart.

And as set out below for a protein counting application (chemical labelling of proteins with nucleic acids):

Protein-forward primer site-label-reverse primer site
Protein1-CAGTCAGTCGCCATCGA-TACCA-GAGTTCGACGGGTAGTC
Protein1-CAGTCAGTCGCCATCGA-ATCAA-GAGTTCGACGGGTAGTC
Protein2-CAGTCAGTCGCCATCGA-ATCAA-GAGTTCGACGGGTAGTC
Protein2-CAGTCAGTCGCCATCGA-GGGCA-GAGTTCGACGGGTAGTC

Even if protein1 and protein2 are labelled with "ATCAA", they could be purified from each other and the labels analyzed thereafter to count how many molecules of each existed in the sample. In this example, the label sequences are flanked by primer sites (forward primer site and reverse primer site) that facilitate amplification of the labels by PCR. The labels can also be used without such sites.

The purpose of the primer site in the above example is to help in amplification of the label sequence by PCR. It is not essential, but it could be included on one or both sides of the label to make the analysis simpler. Without primer sites, one could for example ligate the sequences to a common sequence that helps amplification or sequencing, but may decrease efficiency of analysis (add steps and decreases recovery of labels).

Where the number of molecules of more than one species of interest is to be counted, the number of different labels required to render each molecule of the species of interest distinct is a function of the largest number of identical molecules of a species of interest in the sample. The minimum number of different labels needed to render every molecule of each species of interest in the sample distinct, wherein the largest number of identical molecules of a species of interest in the sample is *n* is *n-1*. For example, where the number of molecules of a first and a second species of interest in a sample are to be counted and the sample contains about 10^5

molecules of the first species of interest and about 10^6 molecules of the second species of interest, about 10^6 labels would be needed for the labelling step to render each molecule in the sample distinct.

As explained above, the methods of the present invention may be used to determine the number of copies of a sequence in a sample. For example, the present invention may be used to determine the number of copies (i.e. repeats) of a defined nucleotide sequence in a chromosome. In accordance with the second development of the first aspect of the invention, set out above, the invention may be used to determine the number of copies of a plurality of sequences in a sample, for example the number of copies of a plurality of sequences in a chromosome.

In a third development of the first aspect of the present invention there is provided a method for determining the number or concentration of molecules of a species of interest in a first sample, and further comprising determining the number or concentration of molecules of a species of interest in a second sample, the method comprising labelling each molecule of the species of interest in the first sample with a first label, wherein each first label is selected from a first group of different labels, to provide a plurality of first conjugates, each first conjugate comprising a first label attached to a molecule of the species of interest, and labelling each molecule of the species of interest in the second sample with a second label, wherein each second label is selected from a second group of different labels, to provide a plurality of second conjugates, each second conjugate comprising a second label attached to a molecule of the species of interest, and determining the number or concentration of molecules of the species of interest in the first sample and the second sample, based on the number of different first conjugates and the number of different second conjugates.

Preferably the first labels are different from the second labels, that is, the labels of the first different group of labels are each different from each of the labels of the second group of different labels.

Preferably, the first group of different labels is a first group of nucleic acids having nucleotide sequences that are different from each other, and the second group of different labels is a second group of nucleic acids having nucleotide sequences that are different from each other and different from each of the nucleic acids in the first group of nucleic acids.

In a preferred embodiment of the third development of the present invention, there is provided a method for determining the number or concentration of molecules of a species of interest in each of a plurality of samples, the method comprising labelling each molecule of the species of

interest with a label, wherein each label is selected from a group of different labels, and wherein each group of different labels used to label each respective sample is different, to provide a plurality of conjugates, each conjugate comprising a label attached to a molecule of the species of interest, and determining the number or concentration of molecules of the species of interest in each sample, based on the number of different conjugates.

The groups of labels may differ from each other in any way that renders a label from one group distinguishable from the label of the other group(s). For example, a different isotope may be used in each different group of labels.

Preferably, the different groups of labels are groups of nucleic acids having nucleotide sequences that are different from each nucleic acid in each other group.

Preferably, different groups of nucleic acid labels differ by nucleic acid length, because nucleic acid length is preserved in amplification. That is, a group of nucleic acids may differ from another group of nucleic acids by the length of nucleic acids in that group. For example a first group of different nucleic acids may consist of nucleic acids each having specific length L_1, and a second group of different nucleic acids may consist of nucleic acids each having specific length L_2, where L_1 and L_2 are different. The lengths of different groups of different nucleic acids may differ in length by 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides. Groups of labels, which groups differ by label length, are especially suitable for methods in which the step of determining the number or concentration of entities of a species of interest comprises mass spectrometry.

Different groups of nucleic acids may differ by nucleotide composition. A group of nucleic acids may differ from another group of nucleic acids by the inclusion of non-natural nucleotides, or mass-tagged nucleotides (i.e. nucleotides containing specific isotopes), in the nucleic acids of that group. A group of nucleic acids may differ from another group of nucleic acids by the modification of the nucleic acids of that group (e.g. CpG methylation). Any kind of modified nucleotide, or nucleic acid, may be used. Preferably, such modified nucleotides or nucleic acids can be digitally copied.

The present invention overcomes disadvantages associated with the methods of the prior art such as microarray-based techniques. Microarray hybridisation can be used to indicate differences in concentration between two or more different samples of a specific mRNA. Since microarrays can be used to simultaneously detect the presence of a large number of mRNAs in a sample, they can be used to indicate differences in concentration between two or more different samples of a large number of mRNAs.

However, microarrays are not suitable for indicating differences in concentration between two or more different mRNAs in a sample. This is because microarrays detect nucleic acids in a sample based on their hybridisation to oligonucleotides tethered to the microarray surface, and hybridisation efficiency varies between different nucleic acids based on their length and sequence structure. Therefore a difference in hybridisation to a microarray between two different mRNAs is not necessarily an indication of the concentration of those mRNAs. Hence, microarrays are suitable for indicating relative differences in concentration of the same mRNA in two different samples (because the length and sequence of those mRNAs is the same) but are not suitable for indicating relative differences in concentration between two or more different mRNAs in a sample. For analogous reasons, microarrays are only suitable for indicating the absolute concentration of a specific mRNA in a sample in certain very limited circumstances (e.g. when the sequence of the specific mRNA is known and hybridisation of the sample to the microarray can be compared that of standard solutions containing known concentrations of that mRNA).

The present invention overcomes these disadvantages of the prior art because it allows the determination of the number or concentration of molecules of a species of interest in a sample, as explained above in relation to the first preferred aspect of the invention. By determining the number or concentration of a molecule of a first species of interest in a sample, and determining the number of concentration of a molecule of a second species of interest in that sample, the numbers or concentrations of molecules of two different species of interest in a sample can be determined, thereby providing information on the absolute and/or relative amounts of the two species in the sample. The second development of the first aspect of the present invention provides the further advantage that the number or concentration of molecules of two or more species of interest in a sample can be determined in a single process, rather than separately determining the number or concentration of molecules of each species of interest in two or more serial or parallel processes.

As mentioned above, microarrays can be used to indicate differences in concentration between two or more different samples of a large number of mRNAs. However, in order for mRNA concentrations between the two or more samples to be compared, two or more microarray hybridisation experiments must be performed.

The present invention provides an advantage over microarray-based methods of determining whether the concentration of a specific mRNA differs between two or more samples. The third development of the first aspect of the present invention allows the number or concentration of molecules of a species of interest (such as a specific mRNA) in two or more samples to be

determined in a single process, rather than performing two or more separate processes and comparing their results.

The second and third developments of the first aspect of the present invention can be combined, in order to determine the number or concentration of molecules of two or more species of interest in two or more samples in a single process.

A second aspect of the present invention relates to tracers.

A tracer is a material used to investigate a reaction or system by tracking progression of a substance in that reaction or system. For example, a tracer may be a material used to monitor progression of a substance in a chemical reaction or industrial process, or in an environmental or biological system. Known tracers include materials such as radioisotopes. Tracers are preferably included in a reaction or system in trace amounts (i.e. relatively very small amounts), so as to minimise interference by the tracer with the reaction or system under investigation. However, small amounts of a tracer material may be difficult to detect, and may be even more difficult to accurately quantify.

As disclosed herein, a population of entities that are different from each other stores information about the number of entities in that population, and this information can be retrieved at different times over a period of time in order to track changes in the number of entities in the population over a period of time. Tracking of the number of entities in the population over time allows the population of entities to function as a tracer.

In a second aspect, the present invention provides a method of tracking a population of entities of a species of interest in a reaction or system, the method comprising providing a plurality of entities of a species of interest that are different from each other, adding the plurality of entities to the reaction or system, and tracking the population of entities by determining the number or concentration of entities of the species of interest in a sample obtained from the reaction or system, based on the number of different entities. The method may comprise generating, or synthesising, the plurality of different entities. The method may comprise a step of generating a plurality of entities of a species of interest that are different from each other.

Optionally, the entities have the structures of the "labels" described elsewhere herein, e.g. they may be polymers such as nucleic acids. They may be synthesised such that they are different from each other, e.g. with different polymer sequences. 20 bp random sequences may be used, or other lengths and molecule types as discussed elsewhere for labels. In the context of the second aspect of the invention, the different entities of the species of interest are effectively "labels", even though they need not be attached to anything, so that a "labelling" step as such need not be performed. For use in the second aspect of the invention, the entities of the

species of interest may simply be mixed with a substance to be monitored, such that the presence of those entities indicates the presence of the substance, and the absolute number or concentration of those entities indicates the amount or concentration of the substance to be monitored, when sampled.

The different entities may indicate flow within the reaction or system. That is, the different entities may merely provide information about the location or movement of a substance within the reaction or system, without their being attached to that substance (e.g. they may mixed with the substance to provided a suspension comprising the entities).

The method may comprise modifying each entity of the species of interest to provide a plurality of differently-modified entities, and determining the number or concentration of entities of the species of interest based on the number of differently-modified entities.

In a first development of the of the second aspect of the invention, there is provided a method of tracking a population of molecules of a species of interest, the method comprising labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates, each conjugate comprising a label attached to a molecule of the species of interest, including the plurality of conjugates in a reaction or system and tracking the molecules of the species of interest in the reaction or system by determining the number or concentration of labelled molecules of the species of interest in a sample obtained from the system, based on the number of different conjugates.

The different conjugates may provide information about the movement or flow of the species of interest in the reaction or system. The species of interest may be a pollutant or contaminant in an environmental or industrial system. The species of interest may be a reactant or substrate in a manufacturing process, such as a drug manufacturing or chemical process.

Two or more different populations of molecules may be tracked in the reaction or system. These different populations of molecules may be added to the reaction or system at different times and/or locations from each other, to provide further information about the reaction or system. The method may comprise providing a first plurality of first molecules that are different from each other, providing a second plurality of second molecules that are different from each other, adding the first and second pluralities of molecules to the reaction or system, and tracking the population of first molecules and tracking the population of second entities by determining the number or concentration of first molecules and second molecules of the species of interest in a sample obtained from the reaction or system, based on the number of different first molecules and the number of different second molecules.

The first and second pluralities of molecules may comprise first and second labels respectively, wherein the first and second labels are different from each other. That is, the labels of the first plurality of molecules are each different the labels of the second plurality of molecules. Preferably, the first labels are first group of nucleic acids, and the labels are second group of nucleic acids having nucleotide sequences that are different from each of the nucleic acids in the first group of nucleic acids.

In the context of the present invention, nucleic acid may be used to track a population of molecules of a species of interest where the molecule of the species of interest is a molecule to which the nucleic acid tracer is attached (i.e. the nucleic acid is a label).

Alternatively, the molecules of the species of interest may be nucleic acid molecules, which are added, or mixed with, a substance to be investigated in a reaction or system. The substance may be a fluid of interest, to which nucleic acid tracer molecules are added, and the flow of the fluid in a system is investigated by obtaining samples of fluid from the system at one or more different time points and/or locations and determining the number or concentration of the nucleic acids in the sample.

In a second development of the second aspect of the invention, there is provided a method of tracking in a reaction or system a population of entities, wherein the entities are nucleic acids, the method comprising providing a plurality of nucleic acids that are different from each other, adding the plurality of nucleic acids to the reaction or system, and tracking the population of nucleic acids by determining the number of nucleic acids in a sample obtained from the reaction or system, based on the number of different nucleic acids.

Nucleic acids to be used as tracers may be nucleic acids having relatively high chemical stability e.g. PNA, LNA. A nucleic acid to be used as a tracer may be in a particle, for example the nucleic acid may be a caged nucleic acid, a coated nucleic acid, a packaged nucleic acid or may be in a virion (i.e. a virus particle).

The method may comprise providing or generating nucleic acids that are different from each other (e.g. 20bp random sequence). The method may comprise adding different populations of entities to a reaction or system, as described above. Two or more different populations of nucleic acids may be synthesised such that each nucleic acid is different (e.g. 20bp random sequence) and such that each population is different (e.g. 20bp predefined sequence, which predefined sequence is different in each group). Different populations of nucleic acids may be added to a reaction or system at different times and/or locations in order to investigate the

reaction or system (e.g. in order to investigate movement or flow of substance within the system). Each population of nucleic acids may be synthesised such that it further comprises a nucleic acid label that identifies or distinguishes the different populations from each other, as described above in relation to different labels for use in labelling molecules of different samples (e.g. different lengths, mass-tagged nucleotides, chemical groups)

An advantage of the methods of the second preferred aspect of the present invention is that the information about the number of molecules in the population, which is stored in the differences between the molecules, is preserved during subsequent processing of the sample. This means that the molecules in the population can be amplified, and/or normalised, in order to improve their detection, and the information about the original number of molecules is preserved. This allows accurate determination of the number or concentration of molecules of the population being tracked in the sample.

In a third development of the second aspect of the present invention, there is provided the use of nucleic acid as a tracer.

Methods and uses of the present invention may be useful in industrial or environmental applications. For example in tracking the composition of complex mixtures, or leak distribution.

In the context of the present invention, a sample contains molecules of a species of interest to be quantified. In accordance with the first aspect of the present invention, the molecules of the species of interest are present in a sample of interest and are subsequently modified, and the number of molecules of the species of interest in the sample is determined as described herein. In accordance with the second aspect of the present invention, molecules of the species of interest are provided that are different from each other and these are subsequently added to a system or reaction of interest, from which a sample is obtained, and then the number or concentration of the molecules of the species of interest in the sample is determined as described herein. In the context of the present invention, the terms "different entities" and "different molecules" may refer to entities/molecules which are different from each other, and may encompass differently-modified entities/molecules (i.e. entities/molecules which differ from each other by virtue of their modification).

In accordance with the methods of the present invention, the molecules of the species of interest in the sample are modified, and the modified population of molecules may be amplified and/or normalised. After the molecules are modified, or after the modified molecules are amplified, or after the modified molecules are normalised, an aliquot of the sample may be obtained, and the number or concentration of molecules of the species of interest in the sample

(i.e. in the original, unprocessed, sample) is determined based on the number of different differently-modified molecules in the aliquot.

Preferably the aliquot is small (contains a low number of molecules) relative to the original sample, library of amplicons, or the normalised library of amplicons. Preferably the aliquot is sufficiently large such that it is possible to detect duplicates of at least one of the labels. The accuracy of the determination of the number or concentration of molecules of the species of interest in the original sample is increased with increasing size of aliquot i.e. the more entities (e.g. conjugates/amplicons) that are detected in order to determine the number of different entities, the more accurate the determination of the number or concentration of molecules of the species of interest. In practice, however, after all labels that conjugated are seen several times, the precision is already extremely high and will increase very little with detection of more entities.

The term determining as used herein means counting, estimating, measuring, quantifying, calculating, deducing, establishing or computing.

The term "number or concentration of molecules of a species of interest" is used hereinto mean either the absolute number, or concentration, or the relative number, or concentration of molecules of a species of interest. Preferably, the method of the present invention is a method for determining the absolute number, amount or concentration of molecules of a species of interest in a sample. The method of the present invention does not necessarily involve determining the absolute number of molecules of a species of interest in a sample, and may involved for example determining the relative number or concentration of molecules of two or more species of interest in a sample. Such information is useful for example in diagnostic applications, as described in more detail below.

Methods of determining the absolute number or concentration of molecules of a species of interest in a sample may comprise determining the relative number of molecules of a species of interest in a sample. For example, determining the absolute number or concentration of molecules of more than one species of interest in a sample, as set out above, provides information on the number or concentration of one species of interest in the sample relative to another species of interest in that sample. Similarly, determining the absolute number or concentration of molecules of a particular species in more than one sample, as set out above, provides information on the number or concentration of molecules of that species of interest in one sample relative to another sample. The method may comprise determining the total number of the genus of molecules to which the species of interest belongs, in which case the method of the invention provides information on the number or concentration of molecules of

the species of interest, relative to the number or concentration of molecules of the genus of interest.

The original number of molecules in a sample, or the number of molecules in the original sample, means number of molecules present in the sample before processing steps that change the number of molecules are performed (e.g. before amplification and/or normalisation).

The number or concentration of molecules of the species of interest is determined based on the number of different entities, e.g. different conjugates, amplicons, or labels, the word "number" here refers to the absolute number of different types of molecule, conjugate or label. For example if, after labelling and amplification of molecules of a species of interest, there were molecules with label A, molecules with label B and molecules with label C, the number of differently labelled (i.e. differently modified) molecules would be three. This could be alternatively expressed as the number of different conjugates being three, or the number of different amplicons is three.

The number of different entities (e.g. different conjugates), may established by detecting the entities and/or the number or distribution of differences between the entities. The number of different entities may be established using standard techniques. For example by sequencing and computational analysis (e.g. string matching, alignment, e.g. Smith Waterman algorithm). Mass spectrometry may also be used to determine the number of different entities. Mass spectrometry may be especially useful for determining the number of different entities in methods which use mass-tagged labels, or labels of differing length.

The number of different entities may be established by sequencing.

The step of determining the number or concentration of entities based on the number of different entities, or differently modified entities, may comprise a step of sequencing a label or conjugate. Sequencing a conjugate comprises sequencing at least part of the label, and may further comprise sequencing at least part of the molecule of the species of interest. Sequencing a conjugate may comprise sequencing a series of monomers that bridge the label and the molecule of the species of interest. Sequencing a conjugate may comprise sequencing the entire conjugate.

A sufficient number of conjugates or amplicons may be detected or sequenced such that a label is observed in duplicate. Preferably a sufficient number of conjugates or amplicons is detected or sequenced such that the majority of labels are observed at least in duplicate i.e. observed at least twice. The more replicates that are observed, the better, e.g. labels can be seen hundreds of times.

A particularly preferred embodiment of the present invention is a method of quantifying the absolute number of DNA-species in a sample without detecting each individual molecule. The method preserves information about the original number of molecules during amplification and/or normalization steps, and can be applied in a multiplexed fashion to count the absolute number of tens of thousands of different DNA species in a sample.

In this particularly preferred embodiment, each individual DNA molecule is labelled with a random DNA sequence label. The labels are designed in such a way that the reaction generates a plurality of DNA molecules in which almost every DNA molecule has a different sequence. As long as the complexity of the plurality of DNA molecules is maintained, they can be (differentially) amplified, normalized and otherwise processed without loss of information about how many DNA molecules were originally present in the sample. This is because making the DNA molecules different from each other during the step of labelling the DNA molecules stores the information about the original number of DNA molecules into a molecular memory consisting of the number of different sequences in the library (Figure 1). Whereas measuring the number of copies generated for each sequence is difficult, counting the number of different sequences is trivial, and this information is not lost during amplification or any other complexity-preserving manipulation of the library.

Sequencing of the library is then used to determine the absolute number of DNA molecules of each species in the original sample (Figure 1). The entire library may be sequenced (following normalisation the library may contain fewer molecules than the original sample). Preferably an aliquot obtained from the library is sequenced, a very small aliquot may be sufficient. With regard to sequences (i.e. conjugates of label sequence and molecule of species of interest sequence), as long as a fraction of duplicates is seen (e.g. 1% duplicates, 99% singlicates), the number of molecules in the sample can be estimated. When all sequences are seen several/tens of times can the number be very accurate. If only singlicates are observed, the lower limit to the number of molecules can be statistically estimated. Such lower limit is larger than the number of labels. If the number of singlicate labels is large and no duplicates are observed, the number of molecules is much larger than the number of observed labels. As noted above, the more replicates that are observed, the better, e.g. hundreds of replicates of the same sequence

The number of molecules in the original sample can be estimated without counting all the molecules, by using sampling statistics. The original number of DNA molecules can be estimated from the sampling distribution of the label sequences for example by assuming

that the number of times a specific label is observed is Poisson distributed. The method to calculate the original number of molecules is well known in mathematics/statistics, this is a sampling problem. For example, the zero truncated Poisson distribution, available for example in the R package described in (Stasinopoulos and Rigby) can be used to derive the original number of molecules from the distribution of sequenced labels that have the same sequence into singlicates, duplicates, triplicates etc. (the zero truncated Poisson distribution is used in this example because the number of un-sequenced labels is unknown). Only a small aliquot of all of the molecules need to be counted in order to statistically estimate the number of molecules in the original sample (Figure 2).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a method in accordance with an embodiment of the invention. DNA-molecules (1) of three different species of interest (light grey, dark grey, black) are labelled with labels selected from a collection of random labels, depicted as different shapes to provide a plurality of conjugates (2). The conjugates are amplified and normalised to provide a normalised library of amplicons (3). After amplification and normalisation, the information about the original number of molecules (1) is preserved in the number of different tags detected by sequencing of an aliquot (4) of the amplified and normalised library. For example, two molecules of a first species of interest are present in the original sample (1, light gray), and two different tags (filled circle, double arrow head) are present in the first species of amplified DNA-molecules sequenced (5, light grey).

Figure 2 shows the results of a simulation of an experiment where labels are used to estimate the original number of mRNA species after normalization of a cDNA library. Simulation results are shown for a total of 82,830 molecules representing eight different cDNA species (frequencies obtained in an actual cDNA normalization experiment, (Zhulidov et al., 2004). Each molecule was given a random 10 bp label (1,048,576 labels), and their frequencies adjusted to correspond to those observed in the amplified and normalized library of Zhulidov et al. Next a random sample (with replacement) of 40,000 molecules was taken from the pool. Ten such simulations were performed and the original number of cDNA molecules prior to normalization (x-axis) was estimated (y-axis, black symbols) by fitting zero-truncated Poisson distribution to the sample label distribution using GAMLSS R package (Stasinopoulos & Rigby, 2007). The number of cDNA sequences for each gene are also shown random sample of labels (grey symbols). Note that accurate estimates (black symbols) can be derived even when the normalization decreases high abundance cDNA (GAPD, gray circles) to a level that is lower than that of medium-abundance cDNA (RPS9, gray diamond).

Figure 3 is a flow chart showing a generic overview of methods in accordance with embodiments of the invention. Groups of molecules are represented by rectangles and processing steps are represented by ovals. A population comprising molecules of a species of interest is processed to provide a sample in which each molecule of the species of interest is different. This processing may comprise a labelling step and/or a bottlenecking step. The processing step may comprise only a labelling step , or only a bottlenecking step, or may comprise a combination of these steps, in either order or simultaneously. The molecules of the sample are then amplified to provide a library of amplicons in which the number of different amplicons indicates the number of molecules of the species of interest in the sample. The library of amplicons may be normalised. The molecules of the species of interest are thus counted by determining the number of different amplicons in the library of amplicons. The number of different amplicons may be determined by observing every amplicon in the library, or by observing every amplicon in an aliquot of the library, or my observing amplicons until at least one duplicate, or a threshold number of replicates, is observed.

Figure 4 is a schematic representation of a virtual karyotyping method in accordance with an embodiment of the invention. The method is used to determine the relative number of two different chromosomes (black and white) present in a specimen (i). A fragmentation step (ii) provides a population of chromosome fragments (iii), in the figure different fragments of the white chromosome are represented by different numbers and different fragments of the black chromosome are represented by different letters. By chance some fragments will be identical, and these are represented by identical numbers/letters. A bottlenecking step (iv) is then performed to provide a sample (v) in which each fragment is different. Fragments are then amplified and assigned to one of the two different chromosomes (vi). The number of "hits" for each respective chromosome (vi) indicates the relative number of each chromosome in the specimen. In the scheme shown, there are six "hits" for the white chromosome, and four "hits" for the black chromosome, indicating that the relative number of white to black chromosomes in the specimen is 6 to 4 , i.e. 3 to 2.

EXAMPLES

Example 1

A method according to an embodiment of the present invention was tested by simulation, using real abundance data from the normalized library described in Zhulidov et al. 2004. Zhulidov et al is concerned with a novel method of normalisation, and standard methods were used to measure cDNA abundance in normalised and non-normalised samples. The normalization severely disrupts the levels of the cDNAs (see Table 1 of Zhulidov et al 2004, and compare Non-normalized to column "AN"), but the numbers or molecules in the original (non-normalised)

sample can be determined using a method of the invention (black symbols in Fig. 2, "estimated number of molecules"). Fig 2 shows that just taking the raw counts, (grey symbols, which show the number of times a given sequence of interest was observed, not considering the labels at all) gives completely erroneous results after normalization.

Frequency data for eight genes were taken from Table 1 of Zhulidov et al. The following simulation was repeated 10 times (each symbol of a specific shape in Figure 2 shows the result obtained for each simulation). An initial pool of 82,830 molecules consisting of the eight DNA-species was created according to the non-normalized frequencies corresponding to approximately 3 million transcripts in total. Each molecule was given a random label from a set of 1,048,576 different labels corresponding a DNA-label of length 10. Labelled molecules were copied so that their frequencies matched normalized amplified frequencies (column AN of Table I of Zhulidov et al). Next a random sample with replacement of size 40,000 molecules was taken from the pool corresponding to sequencing approximately 21 million tags from whole the transcriptome. Each DNA-species was quantified by fitting zero-truncated Poisson distribution to the sample label distribution using GAMLSS R package (Stasinopoulos and Rigby).

Simulation results are shown for counting eight different DNA-species according to the frequencies obtained in a cDNA normalization experiment (Zhulidov et al.). The measured initial frequencies varied from 0.003% (IGF2R) to 1.561% (GAPD) and normalized frequencies from 0.004% (IGF2R) to 0.079% (RPL13a). Ten simulations were performed and the original number of molecules was estimated (black symbols in Figure 2) from a random sample of labels (gray symbols in Figure 2).

In particular the present invention provides methods of monitoring and tracking dilute and/or complex mixtures, as set out below.

Example 2

The present method can be applied to monitoring of mixing reactions and for tracking of composition of complex mixtures for example in industrial or scientific processes.

In particular, the method is effective in analysis of flow patterns and/or the source of leak in cases where multiple input flows converge to a single output flow and the major source(s) of the flow or leak are unknown (for example tracking and monitoring flow patterns in complex microfluidic devices, in analysis of wastewater or farm run-off to streams or rivers) as a very large number of different sequence tags can be introduced to different potential sources, and

the absolute number of tracer-molecules derived from each source detected at the site of interest.

Example 3

For example, in a drug manufacturing process where a small quantity of compound A is added to a large quantity of compound B, and both are further processed, for example, mixed with other substances and/or freeze dried and reconstituted. By adding tracers to A and B, one can amplify a sample of the final drug and accurately determine the relative (and absolute) concentrations of A and B.

Example 4

The present invention further provides a method of karyotyping, which provides a method of diagnosing Down's Syndrome in a fetus carried by a mother, which comprises determining the number of fragments of fetal and maternal DNA in a blood sample obtained from the mother.

The method can be applied to non-invasive prenatal karyotyping. It is well-known that pregnant mothers carry cell-free circulating DNA in their blood, and that 5-20% of that DNA is derived from the fetus (Lo et al., 2007). Accordingly, it is possible to determine the karyotype of the fetus by determining the abundance of cell-free DNA originating from each chromosome (Fan et al, 2008). For example, if the cell-free DNA consists of 95% maternal and 5% fetal DNA, and if the fetus is Trisomy-21 (i.e. has three copies of chromosome 21, leading to Down's syndrome), then the total amount of cell-free DNA from chromosome 21 should exceed that of any other genomic region of the same size by 2.5%. Importantly, then, a very accurate method of determining absolute or relative DNA concentrations is required, and is provided by the present invention.

For example, in order to accurately detect trisomy 21 in maternal cell-free DNA containing 5% fetal DNA, the following procedure can be applied. First, maternal cell-free DNA is extracted. Typically, one can obtain at least 1 ng DNA from 10 mL serum; this is equivalent to approximately 300 genome copies. To reduce the complexity of the sample, it may be digested with restriction enzymes, such that only a subset of chromosomal fragments will be compatible with adapter ligation. For example, enzymes may be selected that result in 10,000 target chromosomal fragments. Second, this DNA is prepared for sequencing, but using adapters containing random sequences, thus ensuring that almost every sample molecule is distinct. Sample preparation for sequencing will typically involve repairing frayed ends, ligating adapters, amplification by PCR and finally adjustment of concentration for sequencing. Finally, chromosomal copy numbers are determined as described above. Similarly, one may determine

the copy number of any desired part of the genome (i.e. not just whole chromosomes, but also chromosomal regions of interest). In order to obtain a reliable quantitative estimate, it is desirable to sequence to saturation, that is until most random sequences have been observed at least twice. Since 10000 fragments from 300 genome copies were prepared for sequencing, at least about 6 million reads are required to obtain an absolute count of the fragments.

Importantly, since an absolute molecule count can be obtained, the disclosed method is almost completely insensitive to amplification artefacts introduced by the varying GC-content of the different chromosomes, by the length of sample molecules or by their secondary structure. To be precise, such bias affects only the total amount of sequence that must be generated to reach saturation, but does not affect the final quantitative molecule count. This is in contrast to prior art methods (e.g. the Quake method) that show strong GC-induced bias, which necessitates a normalization procedure that may only partially eliminate the bias. (Although example discusses absolute molecule count, it would still be effective if only the relative amount in the blood sample of fetal chromosome 21 DNA was determined (i.e. amount of fetal chromosome 21 relative to another fetal chromosome).

Encoding of the concentration information in the number of distinct random sequences thus allows simultaneous measurement of concentrations of biomolecules that are present in very different amounts in the original sample. The method is compatible with sample indexing using separate DNA barcodes, and will have wide applicability in mRNA tag sequencing, ChIP-sequencing, and diagnostic applications such as karyotyping and DNA copy number analysis from small samples.

Example 5

The present invention can be applied to counting and tracking DNA-fragments using spatially ordered or arranged labels. If one or more species of interest is modified with label molecules that are spatially ordered relative to each other, the identity of the labels will carry information about relationships and/or interactions between the molecules of the species of interest.

When sequencing with short reads, the long chromosomal DNA of the target genome must first be cut into shorter fragments, which are then amplified and sequenced. This results in the inevitable loss of information about the order and orientation of the fragments. If DNA is labelled prior to or simultaneously with the fragmentation using an ordered array or an ordered pair of labels, the order of fragments can be tracked and the number of fragments counted.

For example, a recombinase enzyme can be used to bind to and spatially order two identical DNA labels. The reaction of the recombinase with a genomic DNA sequence can then be used to cleave the genomic DNA into two pieces and insert the identical label sequences into both resulting fragments (on both sides of the cleavage point). The left/right relationship of the two adjacent fragments can then be deduced by reading the labels, as the label sequence is added to the 3' end of the left fragment and to the 5' end of the right fragment.

By using recombinases loaded with different label pairs, a 100 kb molecule can be cut into 500 short fragments, that are on average 200 bp long. Each junction is labeled with a distinct pair of 20 bp labels. Then, if every fragment is sequenced individually, including both junction labels, the complete 100 kb molecule can be assembled unambiguously and without using its actual sequence content. Even a 100 kb molecule consisting entirely of As, perhaps with a single T in one arbitrary position, could be assembled correctly.

Statements of invention

1. A method for determining the number or concentration of entities of a species of interest in a sample, the method comprising:

ensuring that the entities of the species of interest in the sample differ from each other, then

amplifying the different entities of the species of interest in the sample to provide a library of amplicons, then

determining the number or concentration of different amplicons, and thereby determining the number or concentration of entities of the species of interest in the sample.

2. A method according to clause 1, comprising:

modifying each entity of the species of interest to provide a plurality of differently-modified entities, and

determining the number or concentration of entities of the species of interest in the sample, based on the number of differently-modified entities.

3. A method according to clause 1 or clause 2, comprising

providing the sample from a population of entities of the species of interest, wherein the population comprises identical and different entities of the species of interest, and the sample size is such that the entities of the species of interest in the sample differ from each other.

4. A method according to any of the preceding clauses, comprising providing the sample from a larger population of entities, by physically or chemically restricting the proportion of entities from the larger population which can be amplified.

5. A method according to any of clauses 2 to 4, comprising modifying a population of entities of the species of interest to provide a plurality of differently-modified entities; and

providing the sample from the population, wherein the entities of the species of interest in the sample are differently modified.

6. A method of tracking a population of entities of a species of interest in a reaction or system, the method comprising:

providing a plurality of entities of a species of interest that are different from each other;

adding the plurality of entities to the reaction or system; and

tracking the population of entities by determining the number or concentration of entities of the species of interest in a sample obtained from the reaction or system, based on the number of different entities.

7. The method according to clause 6, which is a method of tracking a population of entities of a species of interest, the method comprising:

modifying each entity of the species of interest to provide a plurality of differently-modified entities, wherein

the number or concentration of entities of the species of interest is determined based on the number of differently-modified entities.

8. The method according to any one of the preceding clauses, wherein the entities are molecules and the method comprises:

labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates, each conjugate comprising a label attached to a molecule of the species of interest, and

determining the number or concentration of molecules of the species of interest in the sample, based on the number of different conjugates.

9. The method according to clause 8, wherein the method comprises amplifying the differently modified entities by amplifying at least part of the conjugate, which part comprises at least part of the label.

10. The method according to clause 8 or clause 9, comprising labelling a population of entities of the species of interest to provide a population of conjugates, and providing the sample from that population of conjugates, wherein each conjugate in the sample represents an entity of the species of interest.

11. The method according to clause 10, wherein the number of different labels in the group of different labels is smaller than the number of molecules of the species of interest in the population, and

wherein the sample size is such that the conjugates in the sample differ from each other.

12. The method according to any of clauses 8 to 10, wherein the number of different labels in the group of different labels is larger than the number of molecules of the species of interest which are labelled.

13. The method according to any of clauses 8 to 12, wherein the label is a nucleic acid.

14. The method according to clause 13, comprising the step of providing the group of different labels by generating a group of nucleic acids having random sequences.

15. The method according to any one of the preceding clauses, wherein the species of interest is a nucleic acid having a specific nucleotide sequence.

16. The method according to any of the preceding clauses, wherein the number of different amplicons is determined by sequencing the amplicons until each different amplicon has been observed at least in duplicate.

17. The method according to any one of the preceding clauses, wherein the method comprises normalising the plurality of different entities, differently-modified entities, conjugates, or library of amplicons, to provide a normalised plurality of different entities, differently-modified entities or conjugates, or a normalised library of amplicons .

18. The method according to any one of clauses 8 to 17, wherein

the species of interest belongs to a genus of molecules, and wherein

labelling each molecule of the species of interest further comprises labelling each molecule of the genus of molecules in the sample, to provide a population of labelled molecules comprising the plurality of conjugates.

19. The method according to clause 18, wherein the species of interest is a nucleic acid having a specific sequence, and wherein the genus of molecules is a genus of nucleic acids having various sequences.

20. The method according to clause 18 or clause 19, including the additional step of determining the total number of molecules in the sample belonging to the genus of molecules.

21. The method according to any one of clauses 8 to 20, comprising estimating the number of molecules of the species of interest based on the sampling distribution of observed unique and duplicate conjugates in the plurality of conjugates; and/or comprising estimating the total number of molecules in the sample belonging to the genus of molecules, based on the sampling distribution of observed unique and duplicate labels in the population of labelled molecules.

22. The method according to any one of clauses 8 to 21, comprising determining the number or concentration of molecules of both a primary species of interest and a secondary species of interest in the sample, the method comprising:

labelling each molecule of the primary and secondary species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates comprising primary conjugates and secondary conjugates, each primary conjugate

comprising a label attached to a molecule of a primary species of interest, and each secondary conjugate comprising a label attached to a molecule of a secondary species of interest; and

determining the number or concentration of molecules of the primary species of interest in the sample based on the number of different primary conjugates, and determining the number or concentration of molecules of the secondary species of interest in the sample based on the number of different secondary conjugates.

23. The method according to any one of clauses 8 to 22, which is a method for determining the number or concentration of molecules of each of a plurality of particular species of interest in a sample, the method comprising:

labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of particular conjugates, each particular conjugate comprising a label attached to a molecule of a particular species of interest, and

determining the number or concentration of molecules of each particular species of interest in the sample, based on the number of different particular conjugates.

24. The method according to clause 1, wherein the entities are molecules and the method comprises determining the number or concentration of species of interest in a first sample, and further comprising determining the number or concentration of species of interest in a second sample, the method comprising:

labelling each molecule of the species of interest in the first sample with a first label, wherein each first label is selected from a first group of different labels, to provide a plurality of first conjugates, each first conjugate comprising a first label attached to a molecule of the species of interest, and

labelling each molecule of the species of interest in the second sample with a second label, wherein each second label is selected from a second group of different labels, to provide a plurality of second conjugates, each second conjugate comprising a second label attached to a molecule of the species of interest, and

determining the number or concentration of molecules of the species of interest in the first sample and the second sample, based on the number of different first conjugates and the number of different second conjugates.

25. The method according to clause 24, which further includes the features set out in any one of clauses 9 to 23.

26. The method according to clause 6, wherein the method comprises

providing a first plurality of first entities that are different from each other;

providing a second plurality of second entities that are different from each other;

adding the first and second pluralities of entities to the reaction or system, and

tracking the population of first entities and tracking the population of second entities by determining the number or concentration of first entities and second entities of the species of interest in a sample obtained from the reaction or system, based on the number of different first entities and the number of different second entities.

27. The method according to clause 26, wherein

the first plurality of entities is added to the reaction or system at a first time or location, and

the second plurality of entities is added to the reaction or system at a second time or location.

28. The method according to clause 26 or clause 27, which further includes the features set out in any one of clauses 9 to 23.

29. The method according to any one of the preceding clauses, wherein the entities are molecules, and wherein determining the number or concentration of molecules of the species of interest:

(a) does not comprise detecting each individual molecule of the species of interest; or

(b) does not comprise single molecule sequencing; or

(c) does not comprise microarray hybridisation; or

(d) does not comprise sequencing of tags derived from the species of interest.

30. Use of nucleic acid as a tracer.

31. A method of the invention substantially as described herein.

32. A method of diagnosing a chromosomal abnormality in an individual, which method is a virtual karyotyping method for determining the number of two or more different chromosomes represented in a specimen obtained from the individual, the method comprising

providing a population of chromosome fragments from the specimen,

providing a sample from the population of chromosome fragments, wherein the chromosome fragments in the sample differ from each other,

amplifying the different chromosome fragments in the sample to provide a library of amplicons,

determining the number of different amplicons, thereby determining the number of different chromosome fragments in the sample,

identifying the chromosome from which each fragment in the sample was derived, thereby determining the relative number of the different chromosomes represented in the sample.

Further Statements of invention

1. A method for determining the number or concentration of entities of a species of interest in a sample, the method comprising:

modifying each entity of the species of interest to provide a plurality of differently-modified entities, and

determining the number or concentration of entities of the species of interest in the sample, based on the number of differently-modified entities.

2. A method of tracking a population of entities of a species of interest in a reaction or system, the method comprising:

providing a plurality of entities of a species of interest that are different from each other;

adding the plurality of entities to the reaction or system; and

tracking the population of entities by determining the number or concentration of entities of the species of interest in a sample obtained from the reaction or system, based on the number of different entities.

3. The method according to clause 2, which is a method of tracking a population of entities of a species of interest, the method comprising:

modifying each entity of the species of interest to provide a plurality of differently-modified entities, wherein

the number or concentration of entities of the species of interest is determined based on the number of differently-modified entities.

4. The method according to any one of the preceding clauses, wherein the entities are molecules and the method comprises:

labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates, each conjugate comprising a label attached to a molecule of the species of interest, and

determining the number or concentration of molecules of the species of interest in the sample, based on the number of different conjugates.

5. The method according to clause 4, wherein the label is a nucleic acid.

6. The method according to clause 5, comprising the step of providing the group of different labels by generating a group of nucleic acids having random sequences.

7. The method according to any one of the preceding clauses, wherein the species of interest is a nucleic acid having a specific nucleotide sequence.

8. The method according to any one of the preceding clauses, wherein the method comprises the step of amplifying the different entities, or differently modified entities, which optionally comprises the step of amplifying at least part of the conjugate, which part comprises at least part of the label.

9. The method according to clause 8, wherein the amplification step comprises providing a library of amplicons, and wherein the number of entities of the species of interest in the sample is determined based on the number of different amplicons.

10. The method according to any one of clauses 4 to 9, wherein the method comprises normalising the plurality of different entities, differently-modified entities, conjugates, or library of amplicons, to provide a normalised plurality of different entities, differently-modified entities or conjugates, or a normalised library of amplicons .

11. The method according to any one of clauses 4 to 10, wherein
the species of interest belongs to a genus of molecules, and wherein
labelling each molecule of the species of interest further comprises labelling each molecule of the genus of molecules in the sample, to provide a population of labelled molecules comprising the plurality of conjugates.

12. The method according to clause 11, wherein the species of interest is a nucleic acid having a specific sequence, and wherein the genus of molecules is a genus of nucleic acids having various sequences.

13. The method according to clause 11 or clause 12, including the additional step of determining the total number of molecules in the sample belonging to the genus of molecules.

14. The method according to any one of clauses 4 to 13, comprising estimating the number of molecules of the species of interest based on the sampling distribution of observed unique and duplicate conjugates in the plurality of conjugates; and/or comprising estimating the total number of molecules in the sample belonging to the genus of molecules, based on the sampling distribution of observed unique and duplicate labels in the population of labelled molecules.

15. The method according to any one of clauses 4 to 14, comprising determining the number or concentration of molecules of both a primary species of interest and a secondary species of interest in the sample, the method comprising:

labelling each molecule of the primary and secondary species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates comprising primary conjugates and secondary conjugates, each primary conjugate comprising a label attached to a molecule of a primary species of interest, and each secondary conjugate comprising a label attached to a molecule of a secondary species of interest; and

determining the number or concentration of molecules of the primary species of interest in the sample based on the number of different primary conjugates, and determining the number or concentration of molecules of the secondary species of interest in the sample based on the number of different secondary conjugates.

16. The method according to any one of clauses 4 to 15, which is a method for determining the number or concentration of molecules of each of a plurality of particular species of interest in a sample, the method comprising:

labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of particular conjugates, each particular conjugate comprising a label attached to a molecule of a particular species of interest, and

determining the number or concentration of molecules of each particular species of interest in the sample, based on the number of different particular conjugates.

17. The method according to clause 1, wherein the entities are molecules and the method comprises determining the number or concentration of species of interest in a first sample, and further comprising determining the number or concentration of species of interest in a second sample, the method comprising:

labelling each molecule of the species of interest in the first sample with a first label, wherein each first label is selected from a first group of different labels, to provide a plurality of first conjugates, each first conjugate comprising a first label attached to a molecule of the species of interest, and

labelling each molecule of the species of interest in the second sample with a second label, wherein each second label is selected from a second group of different labels, to provide a plurality of second conjugates, each second conjugate comprising a second label attached to a molecule of the species of interest, and

determining the number or concentration of molecules of the species of interest in the first sample and the second sample, based on the number of different first conjugates and the number of different second conjugates.

18. The method according to clause 17, which further includes the features set out in any one of clauses 5 to 16.

19. The method according to clause 2, wherein the method comprises
providing a first plurality of first entities that are different from each other;
providing a second plurality of second entities that are different from each other;
adding the first and second pluralities of entities to the reaction or system, and
tracking the population of first entities and tracking the population of second entities by determining the number or concentration of first entities and second entities of the species of interest in a sample obtained from the reaction or system, based on the number of different first entities and the number of different second entities.

20. The method according to clause 19, wherein
the first plurality of entities is added to the reaction or system at a first time or location, and
the second plurality of entities is added to the reaction or system at a second time or location.

21. The method according to clause 19 or clause 20, which further includes the features set out in any one of clauses 5 to 16.

22. The method according to any one of the preceding clauses, wherein the entities are molecules, and wherein determining the number or concentration of molecules of the species of interest:
(a) does not comprise detecting each individual molecule of the species of interest; or
(b) does not comprise single molecule sequencing; or
(c) does not comprise microarray hybridisation; or
(d) does not comprise sequencing of tags derived from the species of interest.

23. Use of nucleic acid as a tracer.

24. A method of the invention substantially as described herein.

Abbreviations

EST	Expressed sequence tag
L-DNA	the L form of DNA (enantiomer of naturally occuring D-DNA)
LNA	Locked nucleic acid
PNA	Peptide nucleic acid

Bibliography

Coche and Dewez, *Nucleic Acids Res.* 22, 21 4545-4546 (1994)

Cohen, Chang, Boyer, Helling, *Proc Natl Acad Sci* U S A 70, 3240 (1973)

Fan, Blumenfeld, Chitkara, Hudgins, Quake *Proc. Nat. Acad. Sci USA* (2008)

Holland, *J Biol Chem* **277**, 14363 (2002)

Lo and Chiu, *Nature Reviews Genetics,* 8, 71-77 (2007)

Maxam and Gilbert, *Proc Natl Acad Sci* U S A 74, 560 (1977).

Margulies et al., *Nature* 437, 376 (2005)

Nucleic Acids Res 32, e37 (2004).

Okubo et al., *Nat Genet* 2, 173 (1992)

Ozsolak et al., *Nat Methods* 7, 619 (2010).

Patanjali, Parimoo, Weissman, *Proc Natl Acad Sci U S A* 88, 1943 (1991).

Saiki et al., *Science* 230, 1350 (1985).

Sanger, and Coulson, *J Mol Biol* 94, 441 (1975)

Schena, Shalon, Davis, Brown, *Science* 270, 467 (1995)

Shendure et al., *Science* 309, 1728 (2005)

Soares, Bonaldo, Jelene, Su, Lawton and. Efstratiadis, *Proc Natl Acad Sci U S A*, (1994)

Stasinopoulos, and Rigby, *Journal of Statistical Software* 23, 1 (2007).

Velculescu, Zhang, Vogelstein, Kinzler, Science 270, 484 (1995)

Wei et al., *EMBO* J 29, 2147 (2010).

Zhulidov et al., *Nucleic Acids Res.* 32, e37 (2004)

CLAIMS

1. A method for determining the number or concentration of entities of a species of interest in a sample, the method comprising:

ensuring that the entities of the species of interest in the sample differ from each other, then

amplifying the different entities of the species of interest in the sample to provide a library of amplicons, then

determining the number or concentration of different amplicons, and thereby determining the number or concentration of entities of the species of interest in the sample.

2. A method according to claim 1, comprising:

modifying each entity of the species of interest to provide a plurality of differently-modified entities, and

determining the number or concentration of entities of the species of interest in the sample, based on the number of differently-modified entities.

3. A method according to claim 1 or claim 2, comprising

providing the sample from a population of entities of the species of interest, wherein the population comprises identical and different entities of the species of interest, and the sample size is such that the entities of the species of interest in the sample differ from each other.

4. A method according to any of the preceding claims, comprising providing the sample from a larger population of entities, by physically or chemically restricting the proportion of entities from the larger population which can be amplified.

5. A method according to any of claims 2 to 4, comprising modifying a population of entities of the species of interest to provide a plurality of differently-modified entities; and

providing the sample from the population, wherein the entities of the species of interest in the sample are differently modified.

6. A method of tracking a population of entities of a species of interest in a reaction or system, the method comprising:

providing a plurality of entities of a species of interest that are different from each other;

adding the plurality of entities to the reaction or system; and

tracking the population of entities by determining the number or concentration of entities of the species of interest in a sample obtained from the reaction or system, based on the number of different entities.

7. The method according to claim 6, which is a method of tracking a population of entities of a species of interest, the method comprising:

modifying each entity of the species of interest to provide a plurality of differently-modified entities, wherein

the number or concentration of entities of the species of interest is determined based on the number of differently-modified entities.

8. The method according to any one of the preceding claims, wherein the entities are molecules and the method comprises:

labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates, each conjugate comprising a label attached to a molecule of the species of interest, and

determining the number or concentration of molecules of the species of interest in the sample, based on the number of different conjugates.

9. The method according to claim 8, wherein the method comprises amplifying the differently modified entities by amplifying at least part of the conjugate, which part comprises at least part of the label.

10. The method according to claim 8 or claim 9, comprising labelling a population of entities of the species of interest to provide a population of conjugates, and providing the sample from that population of conjugates, wherein each conjugate in the sample represents an entity of the species of interest.

11. The method according to claim 10, wherein the number of different labels in the group of different labels is smaller than the number of molecules of the species of interest in the population, and

wherein the sample size is such that the conjugates in the sample differ from each other.

12. The method according to any of claims 8 to 10, wherein the number of different labels in the group of different labels is larger than the number of molecules of the species of interest which are labelled.

13. The method according to any of claims 8 to 12, wherein the label is a nucleic acid.

14. The method according to claim 13, comprising the step of providing the group of different labels by generating a group of nucleic acids having random sequences.

15. The method according to any one of the preceding claims, wherein the species of interest is a nucleic acid having a specific nucleotide sequence.

16. The method according to any of the preceding claims, wherein the number of different amplicons is determined by sequencing the amplicons until each different amplicon has been observed at least in duplicate.

17. The method according to any one of the preceding claims, wherein the method comprises normalising the plurality of different entities, differently-modified entities, conjugates, or library of amplicons, to provide a normalised plurality of different entities, differently-modified entities or conjugates, or a normalised library of amplicons .

18. The method according to any one of claims 8 to 17, wherein

the species of interest belongs to a genus of molecules, and wherein

labelling each molecule of the species of interest further comprises labelling each molecule of the genus of molecules in the sample, to provide a population of labelled molecules comprising the plurality of conjugates.

19. The method according to claim 18, wherein the species of interest is a nucleic acid having a specific sequence, and wherein the genus of molecules is a genus of nucleic acids having various sequences.

20. The method according to claim 18 or claim 19, including the additional step of determining the total number of molecules in the sample belonging to the genus of molecules.

21. The method according to any one of claims 8 to 20, comprising estimating the number of molecules of the species of interest based on the sampling distribution of observed unique and duplicate conjugates in the plurality of conjugates; and/or comprising estimating the total number of molecules in the sample belonging to the genus of molecules, based on the sampling distribution of observed unique and duplicate labels in the population of labelled molecules.

22. The method according to any one of claims 8 to 21, comprising determining the number or concentration of molecules of both a primary species of interest and a secondary species of interest in the sample, the method comprising:

labelling each molecule of the primary and secondary species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of conjugates comprising primary conjugates and secondary conjugates, each primary conjugate

comprising a label attached to a molecule of a primary species of interest, and each secondary conjugate comprising a label attached to a molecule of a secondary species of interest; and

determining the number or concentration of molecules of the primary species of interest in the sample based on the number of different primary conjugates, and determining the number or concentration of molecules of the secondary species of interest in the sample based on the number of different secondary conjugates.

23. The method according to any one of claims 8 to 22, which is a method for determining the number or concentration of molecules of each of a plurality of particular species of interest in a sample, the method comprising:

labelling each molecule of the species of interest with a label, wherein each label is selected from a group of different labels, to provide a plurality of particular conjugates, each particular conjugate comprising a label attached to a molecule of a particular species of interest, and

determining the number or concentration of molecules of each particular species of interest in the sample, based on the number of different particular conjugates.

24. The method according to claim 1, wherein the entities are molecules and the method comprises determining the number or concentration of species of interest in a first sample, and further comprising determining the number or concentration of species of interest in a second sample, the method comprising:

labelling each molecule of the species of interest in the first sample with a first label, wherein each first label is selected from a first group of different labels, to provide a plurality of first conjugates, each first conjugate comprising a first label attached to a molecule of the species of interest, and

labelling each molecule of the species of interest in the second sample with a second label, wherein each second label is selected from a second group of different labels, to provide a plurality of second conjugates, each second conjugate comprising a second label attached to a molecule of the species of interest, and

determining the number or concentration of molecules of the species of interest in the first sample and the second sample, based on the number of different first conjugates and the number of different second conjugates.

25. The method according to claim 24, which further includes the features set out in any one of claims 9 to 23.

26. The method according to claim 6, wherein the method comprises

providing a first plurality of first entities that are different from each other;

providing a second plurality of second entities that are different from each other;

adding the first and second pluralities of entities to the reaction or system, and

tracking the population of first entities and tracking the population of second entities by determining the number or concentration of first entities and second entities of the species of interest in a sample obtained from the reaction or system, based on the number of different first entities and the number of different second entities.

27. The method according to claim 26, wherein

the first plurality of entities is added to the reaction or system at a first time or location, and

the second plurality of entities is added to the reaction or system at a second time or location.

28. The method according to claim 26 or claim 27, which further includes the features set out in any one of claims 9 to 23.

29. The method according to any one of the preceding claims, wherein the entities are molecules, and wherein determining the number or concentration of molecules of the species of interest:

(a) does not comprise detecting each individual molecule of the species of interest; or

(b) does not comprise single molecule sequencing; or

(c) does not comprise microarray hybridisation; or

(d) does not comprise sequencing of tags derived from the species of interest.

30. Use of nucleic acid as a tracer.

31. A method of the invention substantially as described herein.

32. A method of diagnosing a chromosomal abnormality in an individual, which method is a virtual karyotyping method for determining the number of two or more different chromosomes represented in a specimen obtained from the individual, the method comprising

providing a population of chromosome fragments from the specimen,

providing a sample from the population of chromosome fragments, wherein the chromosome fragments in the sample differ from each other,

amplifying the different chromosome fragments in the sample to provide a library of amplicons,

determining the number of different amplicons, thereby determining the number of different chromosome fragments in the sample,

identifying the chromosome from which each fragment in the sample was derived, thereby determining the relative number of the different chromosomes represented in the sample.

33. The method according to any one of the preceding claims, wherein one or more species of interest is modified with label molecules that are spatially ordered with respect to each other

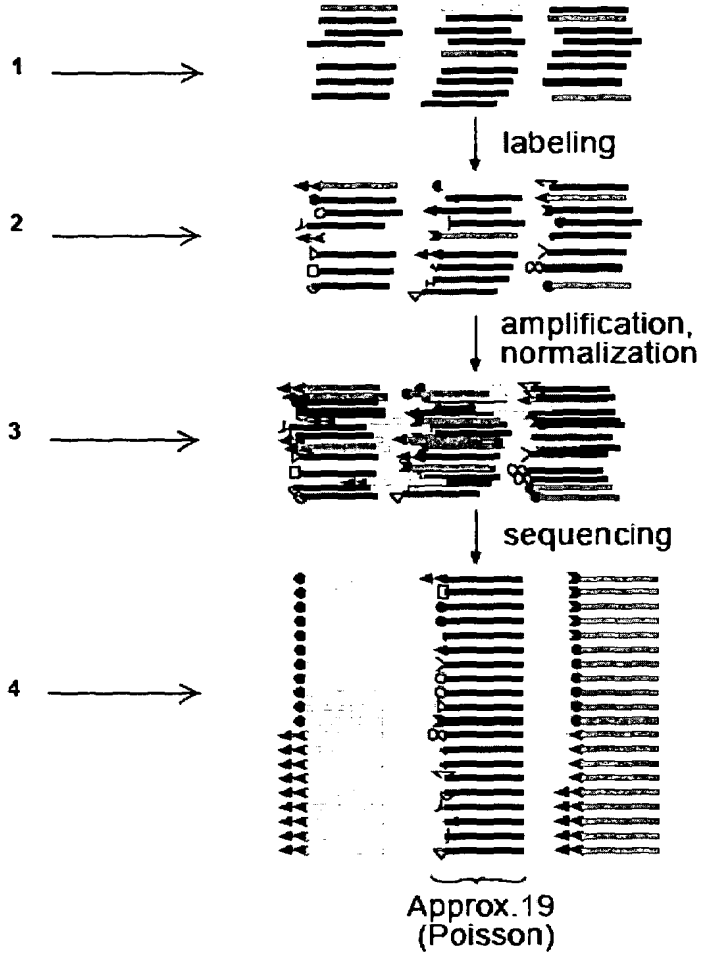

1
labeling
2
amplification,
normalization
3
sequencing
4
Approx.19
(Poisson)

FIGURE 1

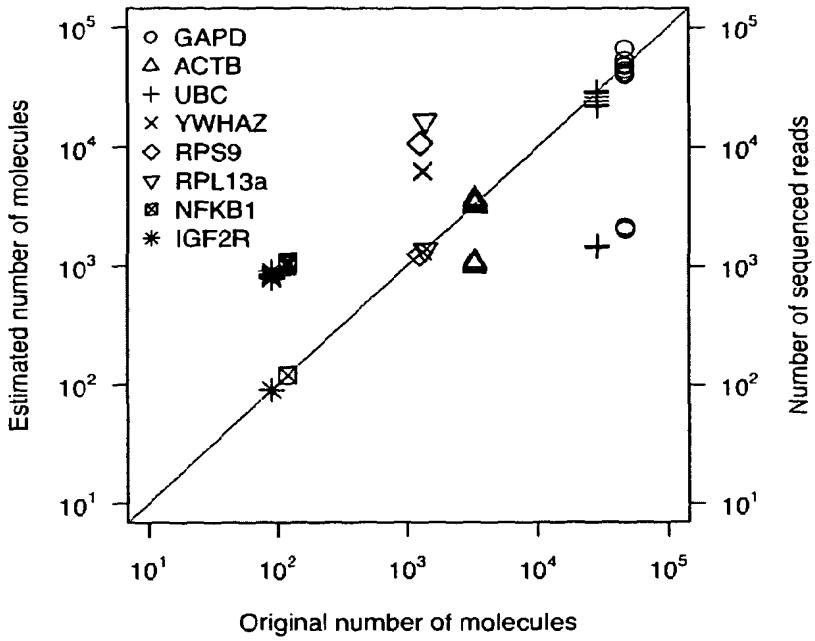

GAPD
ACTB
UBC
YWHAZ
RPS9
RPL13a
NFKB1
IGF2R
Estimated number of molecules
Number of sequenced reads
Original number of molecules
10^1
10^2
10^3
10^4
10^5

FIGURE 2

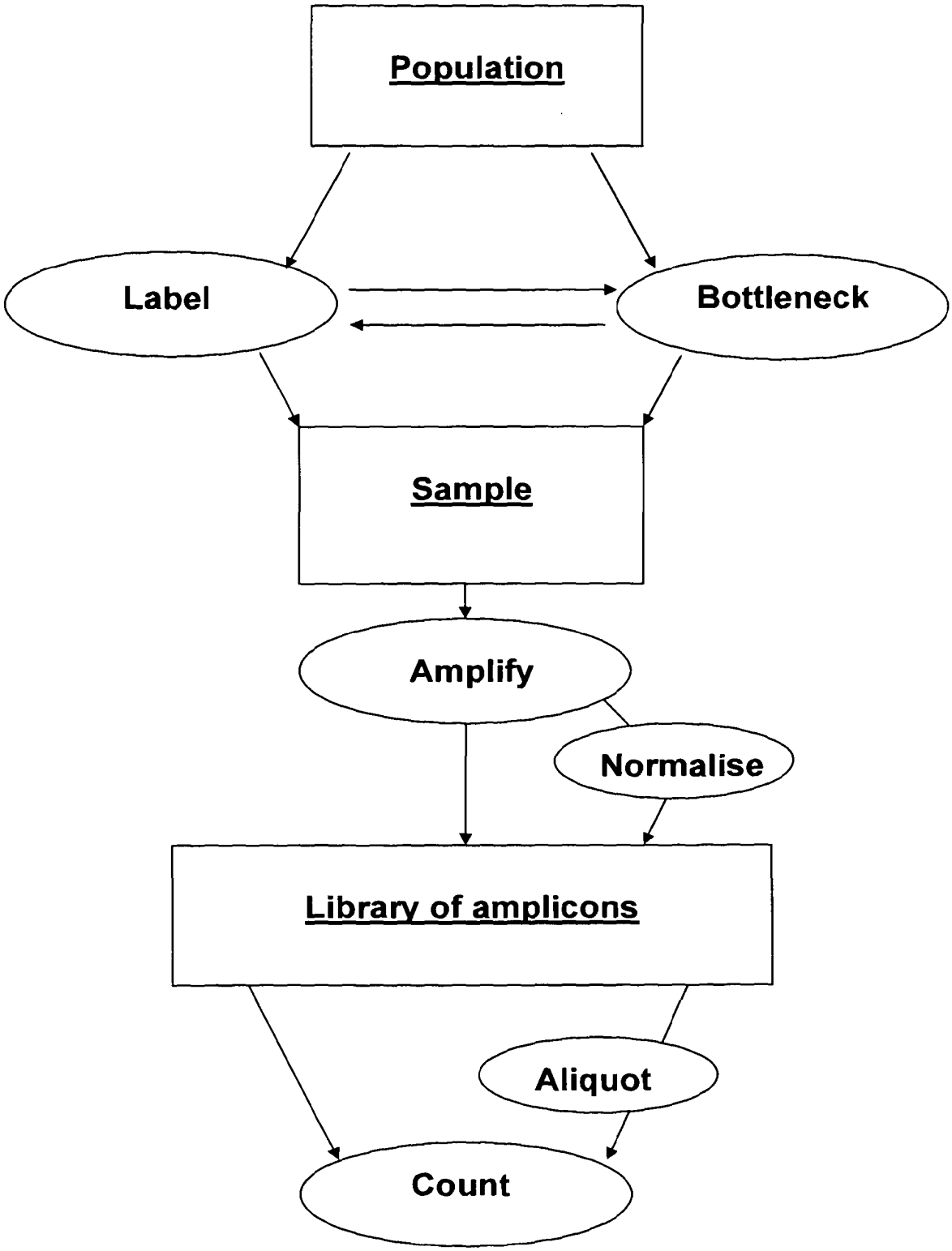

Population
Label
Bottleneck
Sample
Amplify
Normalise
Library of amplicons
Aliquot
Count

FIGURE 3

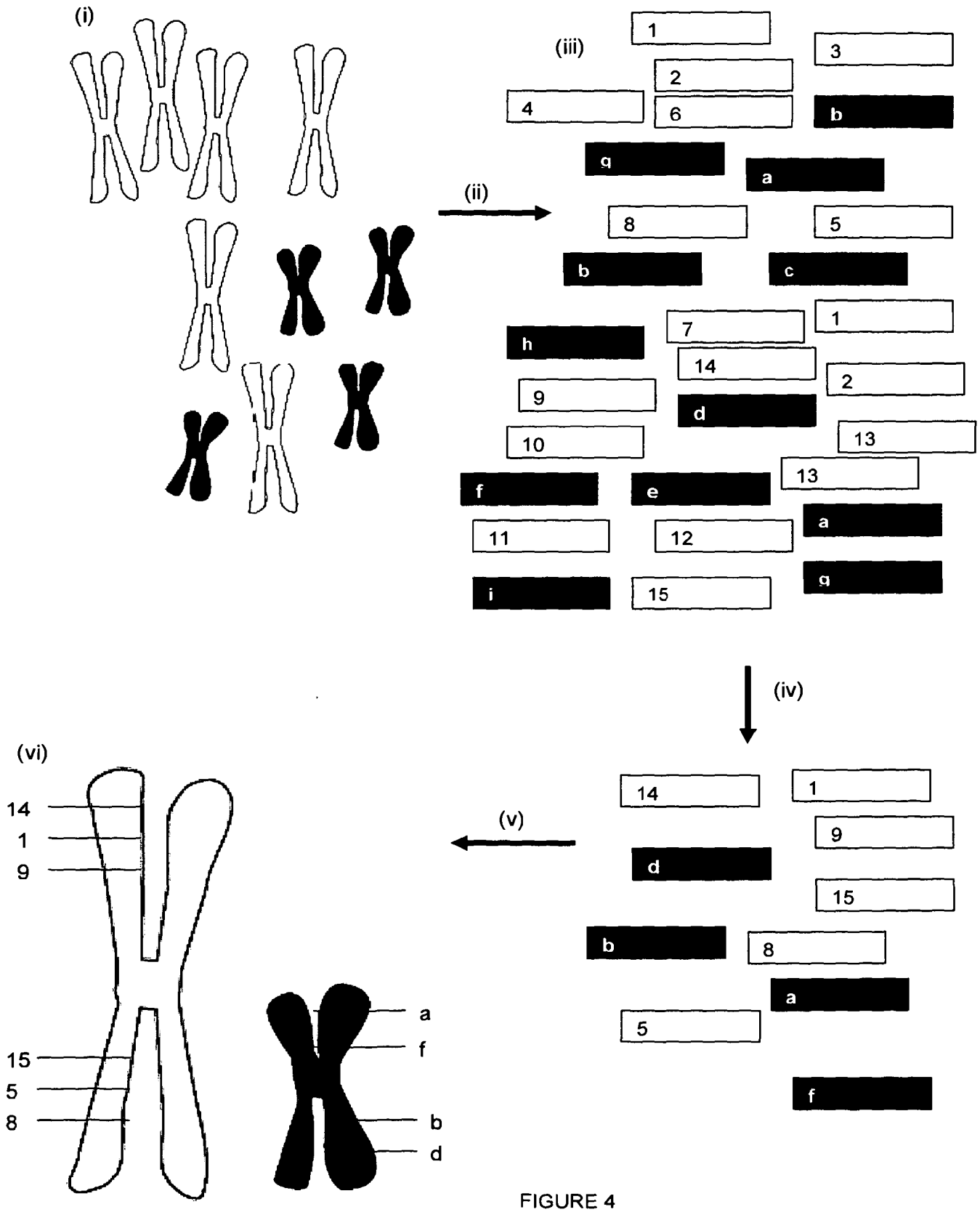

(i)
(ii)
(iii)
(iv)
(v)
(vi)
1
3
2
4
6
b
g
a
8
5
b
c
7
1
h
14
2
9
d
13
10
13
f
e
a
11
12
g
i
15
14
1
9
d
15
b
8
a
5
f
14
1
9
15
5
8
a
f
b
d

FIGURE 4

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 September 2012 (27.09.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/129363 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2012/030039

(22) International Filing Date:
22 March 2012 (22.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/467,037 24 March 2011 (24.03.2011) US
61/583,787 6 January 2012 (06.01.2012) US

(71) Applicant *(for all designated States except US)*: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **XIE, Xiaoliang Sunney** [US/US]; 135 Grant Street, Lexington, Massachusetts 02420 (US). **SHIROGUCHI, Katsuyuki** [JP/US]; 385 Massachusetts Avenue, #50, Arlington, Massachusetts 02474 (US). **SIMS, Peter A.** [US/US]; 872 Massachusetts Avenue, #409, Cambridge, Massachusetts 02139 (US). **JIA, Tony Z.** [US/US]; 12 Oxford Street, #131, Cambridge, Massachusetts 02138 (US).

(74) Agent: **IWANICKI, John P.**; 28 State Street, Suite 1800, Boston, Massachusetts 02109 (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, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

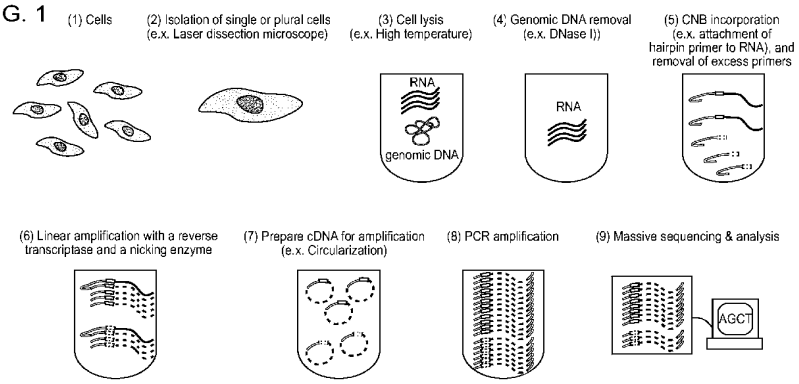
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: SINGLE CELL NUCLEIC ACID DETECTION AND ANALYSIS

FIG. 1



(57) **Abstract**: Methods and compositions for digital profiling of nucleic acid sequences present in a sample are provided.

WO 2012/129363 A2

SINGLE CELL NUCLEIC ACID DETECTION AND ANALYSIS

RELATED APPLICATION DATA

[001] This application claims priority to U.S. Provisional Patent Application No. 61/467,037, filed on March 24, 2011 and U.S. Provisional Patent Application No. 61/583,787, filed on January 6, 2012, each of which is hereby incorporated by reference in their entireties for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

[002] This invention was made with government support under HG005097-01 and 1RC2HG005613-01 from the National Human Genome Research Institute. The Government has certain rights in the invention.

BACKGROUND

Field of the Invention

[003] Embodiments of the present invention relate in general to methods and compositions for determining the expression profile of one or more (e.g., a plurality) nucleic acid sequences in a sample.

Description of Related Art

[004] Although an organism's cells contain largely identical copies of genomic DNA, the RNA expression levels vary widely from cell to cell. RNA expression profiling can be measured. See Pohl and Shih (2004) *Expert Rev. Mol. Diagn.* 4:41 and Wang et al., Nature Reviews 10, 57-63 (2009), WO2007/076128, US2007/0161031 and WO2007/117620. Expression profiling has also been applied to single cells. See Tang et al., *Nature Methods*, 6, 377-382 (2009) and Tang et al., *Cell Stem Cell*, 6, 468-478 (2010). However, many nucleic acids exist at low copy numbers in a cell making detection difficult and a complete expression level profile involves a network containing thousands of genes. Therefore, a technique for genome-wide, digital

quantification of nucleic acid molecules with a high dynamic range and single molecule sensitivity is needed to answer fundamental biological questions.

SUMMARY

[005] Embodiments of the present disclosure are directed to a method of identifying target molecules in a sample, such as a plurality of target molecules from a single cell, using unique barcode sequences. According to one aspect, target molecules include nucleic acid molecules such as DNA or RNA, for example cDNA or mRNA. According to one aspect, a sample is provided including a plurality of nucleic acid molecules, such as DNA or RNA molecules, for example cDNA or mRNA. A nucleic acid molecule in the sample is tagged or labeled with its own unique barcode sequence, i.e. one unique barcode sequence for that particular nucleic acid molecule in the sample, or a combination of two or more barcode sequences providing a unique total barcode sequence for that particular nucleic acid molecule in the sample. The terms unique barcode sequence and unique total barcode sequence are used interchangeably herein. Additional nucleic acid molecules in the sample are also tagged or labeled with their own unique barcode sequences, i.e. one unique barcode sequence for that particular nucleic acid molecule in the sample, or a combination of two or more barcode sequences providing a unique total barcode sequence for that particular nucleic acid molecule in the sample. In this manner, tagged nucleic acid molecules in the sample have different barcode sequences, regardless of whether the individual nucleic acid molecules have the same or a different sequence. A plurality of nucleic acid molecules having the same sequence are referred to herein as having a copy number. According to one aspect, the percentage of individual nucleic acid molecules having their own unique barcode sequence is greater than 80%, greater than 90%, greater than 95%, greater than 96%, greater than 97%, greater than 98%, greater than 99%, greater than 99.1%, greater than 99.2%, greater than 99.3%, greater than 99.4%, greater than 99.5%, greater than 99.6%, greater than 99.7%, greater than 99.8%, or greater than 99.9%. According to one aspect, each nucleic acid molecule, such as each molecule of RNA, such as mRNA, or DNA, such as cDNA in the sample has its own unique barcode sequence, such that 100% of the nucleic acid molecules in the sample have their own unique barcode sequence. The tagged nucleic acid molecules with their

own unique barcode sequences are then amplified in the case of DNA, such as cDNA, or reverse transcribed into corresponding cDNA. According to one aspect, each cDNA includes a barcode sequence unique to the corresponding RNA from which it was transcribed. Each cDNA is then amplified to produce amplicons of the cDNA. Each amplicon includes the barcode sequence from the particular cDNA that was amplified to produce the amplicons. The amplicons are then sequenced whether produced from DNA or RNA and the barcodes are identified. The number of different or unique barcode sequences equates to the number of DNA or cDNA molecules that were amplified and, accordingly, in the case of cDNA, the number of RNA molecules uniquely tagged in the original sample.

[006] According to one aspect, methods and compositions are provided for digital counting of nucleic acids such as RNA and/or DNA with a high dynamic range by using a nucleic acid tag or copy number barcode (CNB) in combination with DNA sequencing, such as a massively parallel sequencing applied to genome-wide RNA expression profiling as described in Wang et al. (2009) *Nat. Rev.* 10:57 hereby incorporated by reference herein in its entirety for all purposes. The methods and compositions described herein reduce or eliminate the amplification bias that typically prevents high sensitivity counting of RNA and/or DNA molecules in a sample.

[007] In an alternate embodiment, a method of identifying the total number of nucleic acid molecules, such as DNA molecules in a sample, such as cDNA, or RNA molecules in a sample, such as a plurality of RNA molecules from a single cell, using unique barcode sequences is provided. According to one aspect, each molecule of DNA or RNA in a sample is tagged or labeled with its own unique barcode sequence, i.e. one unique barcode sequence for the molecule of DNA or RNA in the sample. In this manner, no tagged DNA or RNA in the sample has the same barcode sequence, regardless of whether DNA or RNA molecules are the same or different. This aspect of the present disclosure utilizes a sufficient number of unique candidate barcode sequences relative to the number of DNA or RNA molecules in a sample such that probability ensures that each DNA or RNA molecule will be tagged with a unique barcode sequence. One of skill in the art will recognize a finite probability of two

DNA or RNA molecules being tagged with the same barcode sequence, but the number of candidate barcode sequences is selected such that the chances of this occurring is as infinitesimally small as possible. For RNA, each RNA with its own unique barcode sequence is then reverse transcribed into cDNA. Each cDNA includes a barcode sequence unique to the RNA from which it was transcribed. Each cDNA is then amplified to produce amplicons of the cDNA. Each amplicon includes the barcode sequence from the particular cDNA that was amplified to produce the amplicons. The amplicons are then sequenced and the barcodes are identified. The number of different or unique barcode sequences equates to the number of cDNA molecules that were amplified and, accordingly, the number of RNA molecules present in the original sample. This aspect of the present disclosure provides an elegant method for identifying the total number of RNA molecules in a sample. This aspect also applies to identifying the total number of nucleic acids such as DNA in a sample such as a single cell where one of skill in the art will recognize that reverse transcription steps are not required when the nucleic acid is DNA. For example, cDNA is tagged with its own unique barcode sequence. Each cDNA is then amplified to produce amplicons of the cDNA. Each amplicon includes the barcode sequence from the particular cDNA that was amplified to produce the amplicons. The amplicons are then sequenced and the barcodes are identified. The number of different or unique barcode sequences equates to the number of cDNA molecules that were amplified.

[008] According to aspects of the present disclosure, a method is provided for identifying low copy number nucleic acids in a sample, such as a plurality of DNA and/or RNA molecules from a single cell. According to one aspect, each molecule of RNA in a sample is tagged or labeled with its own unique barcode sequence, i.e. one unique barcode sequence for the molecule of RNA in the sample. Because each molecule of RNA is tagged with its own unique barcode sequence regardless of copy number for a particular RNA sequence, reverse transcription and amplification will reveal RNA of a particular sequence having low copy number, as the sequencing and reading of barcodes is independent of the copy number of the original RNA sequence. This aspect of the present disclosure provides an elegant method for revealing the presence

of low copy number RNA in a cell where the RNA was previously unknown or undiscovered. This aspect also applies to identifying the total number of nucleic acids such as DNA in a sample such as a single cell where one of skill in the art will recognize that reverse transcription steps are not required when the nucleic acid is DNA. For example, cDNA in low copy number in a sample is tagged with its own unique barcode sequence. Each cDNA is then amplified to produce amplicons of the cDNA. Each amplicon includes the barcode sequence from the particular cDNA that was amplified to produce the amplicons. The amplicons are then sequenced and the barcodes are identified. The number of different or unique barcode sequences equates to the number of cDNA molecules that were amplified.

[009] According to certain aspects of the disclosure, a linear pre-amplification method is provided where RNA tagged with its own unique barcode is repeatedly reverse transcribed into multiple copies of cDNA with the unique barcode. The multiple copies of cDNA with the unique barcode may then be amplified to produce the amplicons. The amplicons are then sequenced and the barcodes are identified. As with the other described methods, the number of different or unique barcode sequences ultimately equates to the number of RNA molecules present in the original sample. This aspect of providing multiple copies of cDNA with the unique barcode provides an elegant method for increasing the populace of cDNA molecules corresponding to a particular RNA molecule from one to between about 2 and about 1000 copies or between about 5 to about 100 copies, which improves the efficiency of amplification and detection of the unique barcode sequence, thereby resulting in an accurate count of the number to RNA molecules in a sample. According to certain aspects, linear pre-amplification can be used in a method where DNA is tagged with its own unique barcode. According to this aspect, repeated replication by a DNA polymerase is used for increasing the copy number of a barcoded DNA molecule from between about 2 to about 1000 copies or between about 5 to about 100 copies.

[010] According to certain aspects of the disclosure, a method of making primers personalized to the RNA of a particular cell is provided. According to this aspect, genomic DNA obtained from a cell is fragmented into lengths of between about 5

bases to about 50 bases, between about 10 bases to about 30 bases, or between about 15 bases to about 20 bases. The fragmented genomic DNA is then used as primers for the reverse transcription of RNA from the same species of cell. According to this aspect of the disclosure, the use of a genomic primer pool to reverse transcribe RNA from the same species of cell increases the efficiency of reverse transcription of RNA to cDNA.

[011] According to an aspect of the present disclosure, a method of determining copy number of a nucleic acid molecule in a sample including a plurality of nucleic acid molecules is provided. According to an aspect, the method includes attaching a unique barcode sequence or a unique barcode sequence-primer conjugate or a unique barcode sequence-adapter conjugate to substantially each of the plurality of nucleic acid molecules in the sample to produce a plurality of barcoded nucleic acid molecules, amplifying the plurality of barcoded nucleic acid molecules in the sample to produce amplicons of the plurality of barcoded nucleic acid molecules, sequencing each amplicon to identify an associated nucleic acid sequence and an associated barcode sequence, selecting a first target nucleic acid sequence and determining the number of unique associated barcode sequences for the first target nucleic acid sequence, wherein the number of unique associated barcode sequences is the copy number of the first target nucleic acid sequence. According to an aspect, the nucleic acid molecules are DNA or RNA. According to an aspect, the plurality of barcoded nucleic acid molecules is a plurality of barcoded RNA molecules and further including the steps of reverse transcribing the plurality of barcoded RNA molecules to produce barcoded cDNA molecules and amplifying the barcoded cDNA molecules to produce amplicons of the barcoded cDNA molecules. According to an aspect, a step is provided of repeatedly reverse transcribing the plurality of barcoded RNA molecules to produce linear pre-amplified barcoded cDNA molecules and amplifying the linear pre-amplified barcoded cDNA molecules to produce amplicons of the linear pre-amplified barcoded cDNA molecules. According to an aspect, the step of repeatedly reverse transcribing the plurality of barcoded RNA molecules includes using reverse transcriptase and a nicking enzyme. According to an aspect, the plurality of barcoded nucleic acid molecules is a plurality of barcoded DNA molecules and further

including the steps of repeated replication of the plurality of barcoded DNA molecules to produce a plurality of pre-amplified barcoded DNA molecules and amplifying the plurality of pre-amplified barcoded DNA molecules to produce amplicons of the plurality of pre-amplified barcoded DNA molecules. According to an aspect, the step of repeated replication of the plurality of barcoded DNA molecules includes using DNA polymerase and a nicking enzyme. According to an aspect, the sample is obtained from one or more cells of a first cell type and wherein the primer of the unique barcode sequence-primer conjugate is generated from genomic DNA of the first cell type.

[012] According to an aspect of the present disclosure, a method of counting nucleic acid molecules in a sample including a plurality of nucleic acid molecules is provided. According to an aspect, the method includes attaching a unique barcode sequence or a unique barcode sequence-primer conjugate or a unique barcode sequence-adapter conjugate to substantially each of the plurality of nucleic acid molecules in the sample to produce a plurality of barcoded nucleic acid molecules, amplifying the plurality of barcoded nucleic acid molecules in the sample to produce amplicons of the plurality of barcoded nucleic acid molecules, sequencing each amplicon to identify an associated barcode sequence, and counting the number of unique associated barcode sequences as a measure of the number of nucleic acid molecules in the sample. According to an aspect, the nucleic acid molecules are DNA or RNA. According to an aspect, the plurality of barcoded nucleic acid molecules is a plurality of barcoded RNA molecules and further including the steps of reverse transcribing the plurality of barcoded RNA molecules to produce barcoded cDNA molecules and amplifying the plurality of barcoded cDNA molecules to produce amplicons of the plurality of barcoded cDNA molecules. According to an aspect, a step is provided of repeatedly reverse transcribing the plurality of barcoded RNA molecules to produce linear pre-amplified barcoded cDNA molecules and amplifying the linear pre-amplified barcoded cDNA molecules to produce amplicons of the linear pre-amplified barcoded cDNA molecules. According to an aspect, the step of repeatedly reverse transcribing the plurality of barcoded RNA molecules includes using reverse transcriptase and a nicking enzyme. According to an aspect, the plurality of barcoded nucleic acid

molecules is a plurality of barcoded DNA molecules and further including the steps of repeated replication of the plurality of barcoded DNA molecules to produce a plurality of pre-amplified barcoded DNA molecules and amplifying the plurality of pre-amplified barcoded DNA molecules to produce amplicons of the plurality of pre-amplified barcoded DNA molecules. According to an aspect, the step of repeated replication of the plurality of barcoded DNA molecules includes using DNA polymerase and a nicking enzyme. According to an aspect, the sample is obtained from one or more cells of a first cell type and wherein the primer of the unique barcode sequence-primer conjugate is generated from genomic DNA of the first cell type.

[013] According to an aspect of the present disclosure, a method of determining copy numbers of nucleic acid molecules in a sample is provided. According to an aspect, the method includes attaching a unique barcode sequence or a unique barcode sequence-primer conjugate or a unique barcode sequence-adapter conjugate to substantially each of the nucleic acid molecules in the sample to produce a plurality of barcoded nucleic acid molecules, amplifying the plurality of barcoded nucleic acid molecules in the sample to produce amplicons of the plurality of barcoded nucleic acid molecules, massively parallel sequencing the amplicons of the plurality of barcoded nucleic acid molecules to identify for each amplicon an associated nucleic acid sequence and an associated barcode sequence, and determining the number of unique associated barcode sequences for each nucleic acid sequence in the sample. According to an aspect, the nucleic acid molecules are DNA or RNA. According to an aspect, the plurality of barcoded nucleic acid molecules is a plurality of barcoded RNA molecules and further including the steps of reverse transcribing the plurality of barcoded RNA molecules to produce barcoded cDNA molecules and amplifying the plurality of barcoded cDNA molecules to produce amplicons of the plurality of barcoded cDNA molecules. According to an aspect, a step is provided of repeatedly reverse transcribing the plurality of barcoded RNA molecules to produce linear pre-amplified barcoded cDNA molecules and amplifying the linear pre-amplified barcoded cDNA molecules to produce amplicons of the linear pre-amplified barcoded cDNA molecules. According to an aspect, the step of repeatedly reverse transcribing

the plurality of barcoded RNA molecules includes using reverse transcriptase and a nicking enzyme. According to an aspect, the plurality of barcoded nucleic acid molecules is a plurality of barcoded DNA molecules and further including the steps of repeated replication of the plurality of barcoded DNA molecules to produce a plurality of pre-amplified barcoded DNA molecules and amplifying the plurality of pre-amplified barcoded DNA molecules to produce amplicons of the plurality of pre-amplified barcoded DNA molecules. According to an aspect, the step of repeated replication of the plurality of barcoded DNA molecules includes using DNA polymerase and a nicking enzyme. According to an aspect, the sample is obtained from one or more cells of a first cell type and wherein the primer of the unique barcode sequence-primer conjugate is generated from genomic DNA of the first cell type.

[014] According to one aspect, one or more barcodes may be attached to a nucleic acid such as DNA or RNA. The one or more barcodes may be attached at any location within the nucleic acid. The one or more barcodes may be attached to either end of the nucleic acid. The one or more barcodes may be attached to each end of the nucleic acid. The one or more barcodes may be attached in tandem or in series to the nucleic acid at one or both ends of the nucleic acid or within the nucleic acid.

[015] According to one aspect, the methods described herein may utilize barcodes that are random sequences or systematically designed sequences as are known in the art. Additionally, the methods described herein may utilize barcodes resulting from an optimization protocol, system or method. In such an optimization protocol, system or method, barcodes are designed or selected such that they are not within a certain distance and are sufficient to maintain uniqueness should a barcode sequence be altered during the amplification and sequencing methods or other enzymatic reactions described herein and known to those of skill in the art. For a given barcode length and number of barcodes in a set, the "distance" refers to the number of times a barcode can change a nucleotide before becoming identical to another barcode in the set. For example, for a two member barcode set of AAA and AAT, the distance would be 1 because AAT need change only 1 nucleotide, i.e. the T to an A, to become identical

with AAA. Likewise AAA need change only 1 nucleotide, i.e. the A to a T, to become identical with AAT. For example, if the selected distance were 9, then the members of the optimized set would have a distance greater than 9, as a barcode with a distance of 9, if 9 nucleic acids were changed, would result in creation of a barcode identical to another member of the set. An acceptable distance between barcodes to maintain uniqueness for a given number of alterations is determined by one or more of the particular application for the barcodes, the length of the barcode, the number of barcodes, the amplification error rate, the copy error rate from enzymatic reactions described herein, the sequencing error rate and the number of nucleic acids to be uniquely tagged.

[016] According to an aspect of the present disclosure, barcode sequences useful in the methods described herein are optimized to produce an optimized set of barcodes. The optimized set of barcodes minimizes sequence-dependent bias and/or amplification noise during digital RNA sequencing. The optimized set of barcodes is characterized by a distance of its members such that any particular member may be altered up to and including the selected distance and still maintain uniqueness within the optimized set. If uniqueness were not maintained within the set, then alterations may produce identical barcodes which may lead to a false data regarding the number of nucleic acid molecules in a particular sample. A set of barcodes with members having a predetermined distance allows counting of nucleic acids within a sample with single-copy resolution despite bias from library preparation, sequence-dependent bias and amplification noise. According to one aspect, alterations which may happen during library preparation, amplification and sequencing do not reduce uniqueness of the set of barcodes. According to one aspect, the optimized set of barcodes reduces false original barcode reads that may otherwise result from sequencing errors. According to one aspect, use of the optimized set of barcodes described herein lowers amplification bias ordinarily resulting from such processes like PCR amplification. According to one aspect, one or more optimized barcodes can be ligated to DNA or RNA and amplified with minimal bias and distinguished from one another despite the accumulation of PCR mutations and sequencing errors.

[017] According to one aspect, an optimized barcode set may be characterized as one maintaining sequence uniqueness of its members to the extent of the predetermined distance for the members. According to an aspect, an optimized set of barcodes is provided whereby the optimized barcodes lack significant sequence overlap with other barcodes, lack significant complementarity with other barcodes, lack significant overlap and complementarity with adapter and primer sequences used in library preparation and sequencing, lack significant sequence overlap and complementarity with sequences of the nucleic acids of interest, such as the transcriptome of a cell, lack significantly long homopolymers, lack high GC-content, lack low GC-content, lack possible secondary structures, or lack repetitive sequences.

[018] According to an aspect, a barcode described herein may be attached to two different locations on target nucleic acid, such as at each end of a target nucleic acid. The two barcode sequences independently attached to the nucleic acid along with the target molecule sequence are then determined using methods known to those of skill in the art, such single read sequencing or paired-end sequencing.

[019] According to an aspect, a set of barcodes are used as building blocks to create a barcode attached to a particular nucleic acid. For example, one or more barcodes from the set may be attached to a nucleic acid giving the nucleic acid a unique barcode. For example, two or more barcodes from the set may be attached to a nucleic acid giving the nucleic acid a unique barcode. In this manner, any number of barcodes within the set can be added to a nucleic acid to provide a unique barcode. For example, a first barcode from the set can be added to each nucleic acid in a sample. Then a second barcode from the set can be added to each nucleic acid in a sample. In this manner, the two barcodes combined provide a unique barcode sequence for the nucleic acid. In addition, a third barcode from the set can be added to each nucleic acid in a sample. In addition, a fourth barcode from the set can be added to each nucleic acid in a sample, and so on up to the number of barcodes in the set. In this manner, the barcodes within the set are used as building blocks to create a unique barcode sequence of desired length for nucleic acids within a sample. In this manner, fewer barcode sequences may be included in a set of unique barcodes to create unique

barcode sequences for nucleic acids in a sample. For example, a set of barcodes, such as an optimized set of barcodes described herein, including 145 barcodes with each having a length of 20 nucleotides will produce 145 x 145 = 21,025 unique barcodes if two barcodes from the set are independently attached to a nucleic acid to create a unique barcode. According to certain aspects, attaching two barcodes to a nucleic acid, such as attaching a barcode at each end of a nucleic acid, may increase the overall randomness of barcode sampling because in certain embodiments the two ends of the nucleic acid may be unlikely to have a similar degree of bias. Although the use of two barcode is exemplified, it is to be understood that any number of barcodes from the set may be independently attached to a nucleic acid sequence to create a unique total barcode sequence for the nucleic acid. In this manner, barcodes from the set can be used as individual building blocks to create a unique barcode for each nucleic acid in a sample.

[020] According to one aspect, a method is provided for designing an optimized barcode set by identifying sequences that lack significant sequence overlap with other barcodes, lack significant complementarity with other barcodes, lack significant overlap and complementarity with adapter and primer sequences used in library preparation and sequencing, lack significant sequence overlap and complementarity with sequences of the nucleic acids of interest, such as the transcriptome of a cell, lack significantly long homopolymers, lack high GC-content, lack low GC-content, lack possible secondary structures, or lack repetitive sequences. According to one aspect, a computer is used with commercially available software to design the optimized barcode set which can be created using random or systematic methods known to those of skill in the art by which barcodes meet criteria described herein.

[021] Further features and advantages of certain embodiments of the present disclosure will become more fully apparent in the following description of the embodiments and drawings thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[022] The foregoing and other features and advantages of the present invention will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

[023] ***Figure 1*** schematically depicts RNA sequencing (RNAseq) using copy number barcode.

[024] ***Figure 2*** schematically depicts hairpin primer configurations according to certain embodiments.

[025] ***Figure 3*** schematically depicts linear amplification of cDNA using a nicking enzyme with reverse transcriptase. (1) A DNA primer which includes a double stranded region, copy number barcode (CNB), and a nicking enzyme recognition site is attached to the 3' end of RNA. (2) A reverse transcriptase generates cDNA which includes the complementary sequence of CNB. (3) A nicking enzyme makes a nick. (4) The reverse transcriptase, which has strand displacement activity, starts to generate the cDNA again from the nicking site. (5) The first generated cDNA completely detaches from the template. (6) Multiple cDNAs which have the same CNB (complementary sequence) are generated from the same RNA, meaning that the RNA is linearly amplified such that digital counting is uncompromised.

[026] ***Figures 4A-4C*** depict construction of a genomic primer pool (GPP). (A) One method to make GPP according to certain embodiments. An adapter which has a restriction enzyme recognition site is attached to both ends of fragmented genomic DNA. The enzyme cuts, for example, 20 bp away from the recognition site which is in the genomic sequence. (B) Fragmentation of genomic DNA as a function of time. Samples were loaded on an agarose gel. Length of fragmented DNA became shorter over time. (C) After incubation with a restriction enzyme of PpuEI (cut at 16 bp away), BpmI (16 bp), or MmeI (20 bp) respectively, the samples were loaded on an acrylamide gel. Separated DNA showed expected lengths.

[027] ***Figure 5*** illustrates that the CNB allows digital DNA counting in a single tube. Sequencing of amplified DNA without CNB (above) and with CNB (bottom) are shown. Top panel sequences are set forth as (SEQ ID NO:16). Bottom panel sequences, starting at top and going downward are set forth as (SEQ ID NO:17), (SEQ ID NO:18), (SEQ ID NO:19), (SEQ ID NO:20), (SEQ ID NO:21), (SEQ ID NO:22), (SEQ ID NO:23), (SEQ ID NO:24), respectively.

[028] ***Figure 6*** schematically depicts the principle of digital counting with CNB.

[029] ***Figure 7*** illustrates RNA counting in a single tube by using the CNB primer. Sequencing of amplified cDNA without CNB (above) and with CNB (bottom) are shown. Top panel sequences are set forth as (SEQ ID NO:30). Bottom panel sequences, starting at top and going downward are set forth as (SEQ ID NO:31), (SEQ ID NO:32), (SEQ ID NO:33), (SEQ ID NO:34), (SEQ ID NO:35), (SEQ ID NO:36), (SEQ ID NO:37), (SEQ ID NO:38), respectively.

[030] ***Figure 8*** is a graph depicting the calculated fraction of molecules that are uniquely barcoded as a function of barcode length for four different sample sizes ($N=10^5,10^6,10^7,10^8$) based on the Poisson distribution.

[031] ***Figure 9*** is a graph depicting the results of qPCR on three DNA templates using a common set of primers. The three templates have different amounts of secondary structure, and are therefore amplified during qPCR with different efficiencies. The graph shows both the relative amount of amplification product generated after 20 cycles of qPCR and the corresponding amplification efficiencies for the three templates.

[032] ***Figure 10A*** is an illustration of the digital counting method described herein for cDNA. Figure 10B is an illustration of one aspect of the present disclosure where a cDNA is tagged with one barcode at each end (i.e., two total barcodes) and the sequence is determined using paired-end reading.

[033] ***Figure 11***A-C are graphs quantifying spike-in sequence.

[034] ***Figure 12A-C*** are histograms of unique barcodes from parallel simulation of the theoretical library using optimized barcodes and random barcodes.

[035] ***Figure 13A-B*** are graphs of down-sampling of all spike-in data by a factor of 10 and the resulting digital counts obtained.

[036] ***Figure 14*** is a histogram of the number of bases between the center positions of all pairs of molecules mapped to the same transcription unit that contain the same barcodes for pairs of molecules both mapped to the sense or antisense strand of *E. coli* (light) and also for pairs of molecules mapping to different strands of *E. coli* genome (dark).

[037] ***Figure 15*** is a histogram of the difference in fragment length for all pairs of molecules mapped to the same transcription unit that contain the same barcode for pairs of molecules both mapped to the sense or antisense strand of *E. coli* genome (light) and also for pairs of molecules mapping to different strands of the *E. coli* genome (dark).

[038] ***Figure 16A-D*** are graphs showing digital quantification of the E.coli transcriptome.

[039] ***Figure 17A-D*** are comparisons of uniquely mapped reads per kilobase of each transcription unit or gene per million total uniquely mapped reads (RPKM) and uniquely mapped digital counts per kilobase of each transcription unit or gene per million total uniquely mapped molecules (DPKM) for all detected transcription units.

[040] ***Figure 18A-C*** are graphs depicting reproducibility of digital and conventional quantification of the *E. coli* transcriptome.

[041] ***Figure 19A-C*** are graphs depicting simulated RNA expression quantification.

[042] ***Figure 20*** is a depiction of a barcoding method for digital counting utilizing stochastic labeling (path B) as compared to a conventional dilution method for digital counting (path A).

DETAILED DESCRIPTION

[043] The present invention is based in part on the discovery of methods and compositions for expression profiling of nucleic acid sequences, such as RNA and DNA. In certain exemplary embodiments, methods of individually counting RNA molecules or DNA molecules in a sample are provided. In certain aspects, these methods include the step of attaching a distinguishable or unique barcode, referred to herein as a copy number barcode or CNB, to a respective nucleic acid molecule, such as a DNA molecule in a sample, such as cDNA or such as a RNA molecule in a sample, such as RNA in a cell. According to one aspect, DNA or RNA molecules in a sample receive their own unique barcode. According to one aspect, RNA molecules in the sample with the unique barcodes are then reverse transcribed into cDNA. The cDNA is then amplified and the sequences of the amplified cDNA with the unique barcodes are then determined using methods known to those skilled in the art, such as next generation sequencing methods that simultaneously determine the sequences of the cDNA molecules in a single sample. DNA, such as cDNA in a sample, can similarly be counted using unique barcode sequences.

[044] Terms and symbols of nucleic acid chemistry, biochemistry, genetics, and molecular biology used herein follow those of standard treatises and texts in the field, e.g., Komberg and Baker, *DNA Replication*, Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry*, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, *Human Molecular Genetics*, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); and the like.

[045] As used herein, the term "nucleoside" refers to a molecule having a purine or pyrimidine base covalently linked to a ribose or deoxyribose sugar. Exemplary nucleosides include adenosine, guanosine, cytidine, uridine and thymidine. Additional exemplary nucleosides include inosine, 1-methyl inosine, pseudouridine, 5,6-dihydrouridine, ribothymidine, ^{2}N-methylguanosine and ^{2,2}N,N-dimethylguanosine (also referred to as "rare" nucleosides). The term "nucleotide"

refers to a nucleoside having one or more phosphate groups joined in ester linkages to the sugar moiety. Exemplary nucleotides include nucleoside monophosphates, diphosphates and triphosphates. The terms "polynucleotide," "oligonucleotide" and "nucleic acid molecule" are used interchangeably herein and refer to a polymer of nucleotides, either deoxyribonucleotides or ribonucleotides, of any length joined together by a phosphodiester linkage between 5' and 3' carbon atoms. Polynucleotides can have any three-dimensional structure and can perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment (for example, a probe, primer, EST or SAGE tag), exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A polynucleotide can comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. The term also refers to both double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of this invention that comprises a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double-stranded form. A polynucleotide is composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); thymine (T); and uracil (U) for thymine when the polynucleotide is RNA. Thus, the term polynucleotide sequence is the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

[046] The terms "RNA," "RNA molecule" and "ribonucleic acid molecule" refer to a polymer of ribonucleotides. The terms "DNA," "DNA molecule" and "deoxyribonucleic acid molecule" refer to a polymer of deoxyribonucleotides. DNA and RNA can be synthesized naturally (e.g., by DNA replication or transcription of DNA, respectively). RNA can be post-transcriptionally modified. DNA and RNA can also be chemically synthesized. DNA and RNA can be single-stranded (i.e., ssRNA and ssDNA, respectively) or multi-stranded (e.g., double stranded, i.e.,

dsRNA and dsDNA, respectively). "mRNA" or "messenger RNA" is single-stranded RNA that specifies the amino acid sequence of one or more polypeptide chains. This information is translated during protein synthesis when ribosomes bind to the mRNA.

[047] As used herein, the term "small interfering RNA" ("siRNA") (also referred to in the art as "short interfering RNAs") refers to an RNA (or RNA analog) comprising between about 10-50 nucleotides (or nucleotide analogs) which is capable of directing or mediating RNA interference. In certain exemplary embodiments, an siRNA comprises between about 15-30 nucleotides or nucleotide analogs, between about 16-25 nucleotides (or nucleotide analogs), between about 18-23 nucleotides (or nucleotide analogs), and even between about 19-22 nucleotides (or nucleotide analogs) (e.g., 19, 20, 21 or 22 nucleotides or nucleotide analogs). The term "short" siRNA refers to an siRNA comprising about 21 nucleotides (or nucleotide analogs), for example, 19, 20, 21 or 22 nucleotides. The term "long" siRNA refers to an siRNA comprising about 24-25 nucleotides, for example, 23, 24, 25 or 26 nucleotides. Short siRNAs may, in some instances, include fewer than 19 nucleotides, e.g., 16, 17 or 18 nucleotides, provided that the shorter siRNA retains the ability to mediate RNAi. Likewise, long siRNAs may, in some instances, include more than 26 nucleotides, provided that the longer siRNA retains the ability to mediate RNAi absent further processing, e.g., enzymatic processing, to a short siRNA.

[048] The terms "nucleotide analog," "altered nucleotide" and "modified nucleotide" refer to a non-standard nucleotide, including non-naturally occurring ribonucleotides or deoxyribonucleotides. In certain exemplary embodiments, nucleotide analogs are modified at any position so as to alter certain chemical properties of the nucleotide yet retain the ability of the nucleotide analog to perform its intended function. Examples of positions of the nucleotide which may be derivitized include the 5 position, e.g., 5-(2-amino)propyl uridine, 5-bromo uridine, 5-propyne uridine, 5-propenyl uridine, etc.; the 6 position, e.g., 6-(2-amino) propyl uridine; the 8-position for adenosine and/or guanosines, e.g., 8-bromo guanosine, 8-chloro guanosine, 8-fluoroguanosine, etc. Nucleotide analogs also include deaza nucleotides, e.g., 7-deaza-adenosine; O- and N-modified (e.g., alkylated, e.g., N6-methyl adenosine, or as otherwise known in the art)

nucleotides; and other heterocyclically modified nucleotide analogs such as those described in Herdewijn, Antisense Nucleic Acid Drug Dev., 2000 Aug. 10(4):297-310.

[049] Nucleotide analogs may also comprise modifications to the sugar portion of the nucleotides. For example the 2' OH-group may be replaced by a group selected from H, OR, R, F, Cl, Br, I, SH, SR, NH_2, NHR, NR_2, COOR, or OR, wherein R is substituted or unsubstituted C_1-C_6 alkyl, alkenyl, alkynyl, aryl, etc.Other possible modifications include those described in U.S. Pat. Nos. 5,858,988, and 6,291,438.

[050] The phosphate group of the nucleotide may also be modified, e.g., by substituting one or more of the oxygens of the phosphate group with sulfur (e.g., phosphorothioates), or by making other substitutions which allow the nucleotide to perform its intended function such as described in, for example, Eckstein, Antisense Nucleic Acid Drug Dev. 2000 Apr. 10(2):117-21, Rusckowski et al. Antisense Nucleic Acid Drug Dev. 2000 Oct. 10(5):333-45, Stein, Antisense Nucleic Acid Drug Dev. 2001 Oct. 11(5): 317-25, Vorobjev et al. Antisense Nucleic Acid Drug Dev. 2001 Apr. 11(2):77-85, and U.S. Pat. No. 5,684,143. Certain of the above-referenced modifications (e.g., phosphate group modifications) decrease the rate of hydrolysis of, for example, polynucleotides comprising said analogs *in vivo* or *in vitro*.

[051] As used herein, the term "isolated RNA" (e.g., "isolated mRNA") refers to RNA molecules which are substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[052] The term "*in vitro*" has its art recognized meaning, e.g., involving purified reagents or extracts, e.g., cell extracts. The term "*in vivo*" also has its art recognized meaning, e.g., involving living cells, e.g., immortalized cells, primary cells, cell lines, and/or cells in an organism.

[053] As used herein, the terms "complementary" and "complementarity" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the

sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be partial or total. Partial complementarity occurs when one or more nucleic acid bases is not matched according to the base pairing rules. Total or complete complementarity between nucleic acids occurs when each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[054] The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (i.e., partial identity) or complete homology (i.e., complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (i.e., an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

[055] When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency. When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can

hybridize to the single-stranded nucleic acid sequence under conditions of low stringency.

[056] The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence," "sequence identity," "percentage of sequence identity" and "substantial identity." A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA sequence given in a sequence listing or may comprise a complete gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window," as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (Smith and Waterman (1981) *Adv. Appl. Math.* 2:482) by the homology alignment algorithm of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970)), by the search for similarity method of Pearson and Lipman (*Proc. Natl. Acad. Sci. USA* 85:2444 (1988)]), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection, and the best

alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

[057] The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of the full-length sequences of the compositions claimed in the present invention.

[058] The term "hybridization" refers to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

[059] The term "T_m" refers to the melting temperature of a nucleic acid. The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41$ (% G + C), when a nucleic acid is in aqueous solution at 1 M NaCl (See, e.g., Anderson and Young, Quantitative Filter Hybridization, in *Nucleic Acid Hybridization* (1985)). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m.

[060] The term "stringency" refers to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted.

[061] "Low stringency conditions," when used in reference to nucleic acid hybridization, comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5x SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4(H_2O)$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5x Denhardt's reagent (50x Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)) and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5x SSPE, 0.1% SDS at 42 °C when a probe of about 500 nucleotides in length is employed.

[062] "Medium stringency conditions," when used in reference to nucleic acid hybridization, comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5x SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4(H_2O)$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5x Denhardt's reagent and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0x SSPE, 1.0% SDS at 42 °C when a probe of about 500 nucleotides in length is employed.

[063] "High stringency conditions," when used in reference to nucleic acid hybridization, comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5x SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4(H_2O)$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5x Denhardt's reagent and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1x SSPE, 1.0% SDS at 42 °C when a probe of about 500 nucleotides in length is employed.

[064] It is well known that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target molecule (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

[065] In certain exemplary embodiments, and with reference to Figure 1, cells are identified and then a single cell or a plurality of cells are isolated. Cells within the scope of the present disclosure include any type of cell where understanding the RNA content is considered by those of skill in the art to be useful. A cell according to the present disclosure includes a hepatocyte, oocyte, embryo, stem cell, iPS cell, ES cell, neuron, erythrocyte, melanocyte, astrocyte, germ cell, oligodendrocyte, kidney cell, leukocyte, thrombocyte, epithelial cell, adipocyte or fibroblast and the like. According to one aspect, the methods of the present invention are practiced with the cellular RNA from a single cell. However, according to certain aspects, the cellular RNA content from a plurality of cells may be used. A plurality of cells includes from about 2 to about 1,000,000 cells, about 2 to about 10 cells, about 2 to about 100 cells, about 2 to about 1,000 cells, about 2 to about 10,000 cells, or about 2 to about 100,000 cells.

[066] As used herein, a "single cell" refers to one cell. Single cells useful in the methods described herein can be obtained from a tissue of interest, or from a biopsy, blood sample, or cell culture. Additionally, cells from specific organs, tissues, tumors, neoplasms, or the like can be obtained and used in the methods described herein. Furthermore, in general, cells from any population can be used in the methods, such as a population of prokaryotic or eukaryotic single celled organisms including bacteria or yeast. In some aspects of the invention, the method of preparing a collection of cDNA, i.e. a cDNA library, can include the step of obtaining single cells. A single cell suspension can be obtained using standard methods known in the art including, for example, enzymatically using trypsin or papain to digest proteins connecting cells in tissue samples or releasing adherent cells in culture, or mechanically separating cells in a sample. Single cells can be placed in any suitable reaction vessel in which single cells can be treated individually. For example a 96-well plate, such that each single cell is placed in a single well.

[067] Methods for manipulating single cells are known in the art and include fluorescence activated cell sorting (FACS), micromanipulation and the use of semi-automated cell pickers (e.g. the Quixell™ cell transfer system from Stoelting Co.). Individual cells can, for example, be individually selected based on features detectable by microscopic observation, such as location, morphology, or reporter gene expression.

[068] Aspects of the present disclosure include methods for identifying the RNA content and copy number for certain RNA sequences in a cell. Using the methods disclosed herein, the RNA content of cells can be compared at various stages or times and the absolute amount of RNA or the relative amount of RNA can be used to diagnose or treat certain diseases, disorders or conditions. Diseases, disorders or conditions within the scope of the present disclosure include cancer, Alzheimer's disease, Parkinson's disease, hepatitis, muscular dystrophy, psoriasis, tuberculosis, lysosome disease, ulcerative colitis, and Ehlers-Danlos Syndrome, cystic fibrosis, diabetes, hemophilia, sickle cell anemia, HIV, autoimmune diseases, Huntington's disease, ALS, and Shy-Drage syndrome.

[069] Once a desired cell has been identified, the cell is lysed to release cellular contents including DNA and RNA, such as mRNA, using methods known to those of skill in the art. The cellular contents are contained within a vessel. In some aspects of the invention, cellular contents, such as mRNA, can be released from the cells by lysing the cells. Lysis can be achieved by, for example, heating the cells, or by the use of detergents or other chemical methods, or by a combination of these. However, any suitable lysis method known in the art can be used. A mild lysis procedure can advantageously be used to prevent the release of nuclear chromatin, thereby avoiding genomic contamination of the cDNA library, and to minimise degradation of mRNA. For example, heating the cells at 72°C for 2 minutes in the presence of Tween-20 is sufficient to lyse the cells while resulting in no detectable genomic contamination from nuclear chromatin. Alternatively, cells can be heated to 65°C for 10 minutes in water (Esumi et al., *Neurosci Res* 60(4):439-51 (2008)); or 70°C for 90 seconds in PCR buffer II (Applied Biosystems) supplemented with 0.5% NP-40 (Kurimoto et al., *Nucleic Acids Res* 34(5):e42 (2006)); or lysis can be achieved with a protease such as Proteinase K or by the use of chaotropic salts such as guanidine isothiocyanate (U.S. Publication No. 2007/0281313).

[070] Nucleic acids from a cell such as DNA or RNA are isolated using methods known to those of skill in the art. Such methods include removing or otherwise separating genomic DNA and other cellular constituents from RNA. Methods of removing or separating genomic DNA from RNA include the use of DNase I.

[071] Synthesis of cDNA from mRNA in the methods described herein can be performed directly on cell lysates, such that a reaction mix for reverse transcription is added directly to cell lysates. Alternatively, mRNA can be purified after its release from cells. This can help to reduce mitochondrial and ribosomal contamination. mRNA purification can be achieved by any method known in the art, for example, by binding the mRNA to a solid phase. Commonly used purification methods include paramagnetic beads (e.g. Dynabeads). Alternatively, specific contaminants, such as ribosomal RNA can be selectively removed using affinity purification.

[072] The nucleic acids, such as DNA or RNA, are then combined in a vessel with primers. The primers include unique barcode sequences and the primers attached to nucleic acid molecules. According to one aspect, nucleic acid molecules within the vessel each have a unique barcode sequence. After primers have attached to the nucleic acid molecules, excess primers are removed from the vessel.

[073] According to one aspect, unique barcode sequences may be attached to each nucleic acid, i.e. DNA or RNA in a sample. Then adapters and or primers or other reagents known to those of skill in the art may be used as desired to reverse transcribe or amplify the nucleic acid with the unique barcode sequence, as the case may be.

[074] cDNA is typically synthesized from mRNA by reverse transcription. Methods for synthesizing cDNA from small amounts of mRNA, including from single cells, have previously been described (see Kurimoto et al., *Nucleic Acids Res* 34(5):e42 (2006); Kurimoto et al., Nat Protoc 2(3):739-52 (2007); and Esumi et al., *Neurosci Res* 60(4):439-51 (2008)). In order to generate an amplifiable cDNA, these methods introduce a primer annealing sequence at both ends of each cDNA molecule in such a way that the cDNA library can be amplified using a single primer. The Kurimoto method uses a polymerase to add a 3' poly-A tail to the cDNA strand, which can then be amplified using a universal oligo-T primer. In contrast, the Esumi method uses a template switching method to introduce an arbitrary sequence at the 3' end of the cDNA, which is designed to be reverse complementary to the 3' tail of the cDNA synthesis primer. Again, the cDNA library can be amplified by a single PCR primer. Single-primer PCR exploits the PCR suppression effect to reduce the amplification of short contaminating amplicons and primer-dimers (Dai et al., *J Biotechnol* 128(3):435-43 (2007)). As the two ends of each amplicon are complementary, short amplicons will form stable hairpins, which are poor templates for PCR. This reduces the amount of truncated cDNA and improves the yield of longer cDNA molecules.

[075] In some aspects of the invention, the synthesis of the first strand of the cDNA can be directed by a cDNA synthesis primer (CDS) that includes an RNA complementary sequence (RCS). In some aspects of the invention, the RCS is at least partially

complementary to one or more mRNA in an individual mRNA sample. This allows the primer, which is typically an oligonucleotide, to hybridize to at least some mRNA in an individual mRNA sample to direct cDNA synthesis using the mRNA as a template. The RCS can comprise oligo (dT), or be gene family-specific, such as a sequence of nucleic acids present in all or a majority related genes, or can be composed of a random sequence, such as random hexamers. To avoid the CDS priming on itself and thus generating undesired side products, a non-self-complementary semi- random sequence can be used. For example, one letter of the genetic code can be excluded, or a more complex design can be used while restricting the CDS to be non-self-complementary.

[076] The RCS can also be at least partially complementary to a portion of the first strand of cDNA, such that it is able to direct the synthesis of a second strand of cDNA using the first strand of the cDNA as a template. Thus, following first strand synthesis, an RNase enzyme (e.g. an enzyme having RNaseH activity) can be added after synthesis of the first strand of cDNA to degrade the RNA strand and to permit the CDS to anneal again on the first strand to direct the synthesis of a second strand of cDNA. For example, the RCS could comprise random hexamers, or a non-self complementary semi-random sequence (which minimizes self- annealing of the CDS).

[077] A template-switching oligonucleotide (TSO) that includes a portion which is at least partially complementary to a portion of the 3' end of the first strand of cDNA can be added to each individual mRNA sample in the methods described herein. Such a template switching method is described in (Esumi et al, Neurosci Res 60(4):439-51 (2008)) and allows full length cDNA comprising the complete 5' end of the mRNA to be synthesized. As the terminal transferase activity of reverse transcriptase typically causes 2-5 cytosines to be incorporated at the 3' end of the first strand of cDNA synthesized from mRNA, the first strand of cDNA can include a plurality of cytosines, or cytosine analogues that base pair with guanosine, at its 3' end (see US 5, 962,272). In one aspect of the invention, the first strand of cDNA can include a 3' portion comprising at least 2, at least 3, at least 4, at least 5 or 2, 3, 4, or 5 cytosines or

cytosine analogues that base pair with guanosine. A non-limiting example of a cytosine analogue that base pairs with guanosine is 5-aminoallyl-2'-deoxycytidine.

[078] In one aspect of the invention, the TSO can include a 3' portion comprising a plurality of guanosines or guanosine analogues that base pair with cytosine. Non-limiting examples of guanosines or guanosine analogues useful in the methods described herein include, but are not limited to, deoxyriboguanosine, riboguanosine, locked nucleic acid-guanosine, and peptide nucleic acid-guanosine. The guanosines can be ribonucleosides or locked nucleic acid monomers.

[079] A locked nucleic acid (LNA) is a modified RNA nucleotide. The ribose moiety of an LNA nucleotide is modified with an extra bridge connecting the 2' oxygen and 4' carbon. The bridge "locks" the ribose in the 3'-endo (North) conformation. Some of the advantages of using LNAs in the methods of the invention include increasing the thermal stability of duplexes, increased target specificity and resistance from exo- and endonucleases.

[080] A peptide nucleic acid (PNA) is an artificially synthesized polymer similar to DNA or RNA, wherein the backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds. The backbone of a PNA is substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This provides two non-limiting advantages. First, the PNA backbone exhibits improved hybridization kinetics. Secondly, PNAs have larger changes in the melting temperature (Tm) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4°C drop in Tm for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9°C. This can provide for better sequence discrimination. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration.

[081] A nucleic acid useful in the invention can contain a non-natural sugar moiety in the backbone. Exemplary sugar modifications include but are not limited to 2'

modifications such as addition of halogen, alkyl substituted alkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, SO_2CH_3, OSO_2, SO_3, CH_3, ONO_2, NO_2, N_3, NH_2, substituted silyl, and the like. Similar modifications can also be made at other positions on the sugar, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Nucleic acids, nucleoside analogs or nucleotide analogs having sugar modifications can be further modified to include a reversible blocking group, peptide linked label or both. In those embodiments where the above-described 2' modifications are present, the base can have a peptide linked label.

[082] A nucleic acid used in the invention can also include native or non-native bases. In this regard a native deoxyribonucleic acid can have one or more bases selected from the group consisting of adenine, thymine, cytosine or guanine and a ribonucleic acid can have one or more bases selected from the group consisting of uracil, adenine, cytosine or guanine. Exemplary non-native bases that can be included in a nucleic acid, whether having a native backbone or analog structure, include, without limitation, inosine, xathanine, hypoxathanine, isocytosine, isoguanine, 5-methylcytosine, 5-hydroxymethyl cytosine, 2-aminoadenine, 6-methyl adenine, 6-methyl guanine, 2-propyl guanine, 2-propyl adenine, 2-thioLiracil, 2-thiothymine, 2-thiocytosine, 15 -halouracil, 15 -halocytosine, 5-propynyl uracil, 5-propynyl cytosine, 6-azo uracil, 6-azo cytosine, 6-azo thymine, 5-uracil, 4-thiouracil, 8-halo adenine or guanine, 8- amino adenine or guanine, 8-thiol adenine or guanine, 8-thioalkyl adenine or guanine, 8- hydroxyl adenine or guanine, 5-halo substituted uracil or cytosine, 7-methylguanine, 7- methyladenine, 8-azaguanine, 8-azaadenine, 7-deazaguanine, 7-deazaadenine, 3-deazaguanine, 3-deazaadenine or the like. A particular embodiment can utilize isocytosine and isoguanine in a nucleic acid in order to reduce non-specific hybridization, as generally described in U.S. Pat. No.5,681,702.

[083] A non-native base used in a nucleic acid of the invention can have universal base pairing activity, wherein it is capable of base pairing with any other naturally occurring base. Exemplary bases having universal base pairing activity include 3-nitropyrrole and 5- nitroindole. Other bases that can be used include those that have

base pairing activity with a subset of the naturally occurring bases such as inosine, which basepairs with cytosine, adenine or uracil.

[084] In one aspect of the invention, the TSO can include a 3' portion including at least 2, at least 3, at least 4, at least 5, or 2, 3, 4, or 5, or 2-5 guanosines, or guanosine analogues that base pair with cytosine. The presence of a plurality of guanosines (or guanosine analogues that base pair with cytosine) allows the TSO to anneal transiently to the exposed cytosines at the 3' end of the first strand of cDNA. This causes the reverse transcriptase to switch template and continue to synthesis a strand complementary to the TSO. In one aspect of the invention, the 3' end of the TSO can be blocked, for example by a 3' phosphate group, to prevent the TSO from functioning as a primer during cDNA synthesis.

[085] In one aspect of the invention, the mRNA is released from the cells by cell lysis. If the lysis is achieved partially by heating, then the CDS and/or the TSO can be added to each individual mRNA sample during cell lysis, as this will aid hybridization of the oligonucleotides. In some aspects, reverse transcriptase can be added after cell lysis to avoid denaturation of the enzyme.

[086] In some aspects of the invention, a tag can be incorporated into the cDNA during its synthesis. For example, the CDS and/or the TSO can include a tag, such as a particular nucleotide sequence, which can be at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15 or at least 20 nucleotides in length. For example, the tag can be a nucleotide sequence of 4-20 nucleotides in length, e.g. 4, 5, 6, 7, 8, 9, 10, 15 or 20 nucleotides in length. As the tag is present in the CDS and/or the TSO it will be incorporated into the cDNA during its synthesis and can therefore act as a "'barcode" to identify the cDNA. Both the CDS and the TSO can include a tag. The CDS and the TSO can each include a different tag such that the tagged cDNA sample comprises a combination of tags. Each cDNA sample generated by the above method can have a distinct tag, or a distinct combination of tags, such that once the tagged cDNA samples have been pooled, the tag can be used to identify from which single cell each cDNA sample originated. Thus, each cDNA sample can be linked to a

single cell, even after the tagged cDNA samples have been pooled in the methods described herein.

[087] Before the tagged cDNA samples are pooled, synthesis of cDNA can be stopped, for example by removing or inactivating the reverse transcriptase. This prevents cDNA synthesis by reverse transcription from continuing in the pooled samples. The tagged cDNA samples can optionally be purified before amplification, ether before or after they are pooled.

[088] As used herein, the term "barcode" refers to a unique oligonucleotide sequence that allows a corresponding nucleic acid base and/or nucleic acid sequence to be identified. In certain aspects, the nucleic acid base and/or nucleic acid sequence is located at a specific position on a larger polynucleotide sequence (e.g., a polynucleotide covalently attached to a bead). In certain embodiments, barcodes can each have a length within a range of from 4 to 36 nucleotides, or from 6 to 30 nucleotides, or from 8 to 20 nucleotides. In certain aspects, the melting temperatures of barcodes within a set are within 10 °C of one another, within 5 °C of one another, or within 2 °C of one another. In other aspects, barcodes are members of a minimally cross-hybridizing set. That is, the nucleotide sequence of each member of such a set is sufficiently different from that of every other member of the set that no member can form a stable duplex with the complement of any other member under stringent hybridization conditions. In one aspect, the nucleotide sequence of each member of a minimally cross-hybridizing set differs from those of every other member by at least two nucleotides. Barcode technologies are known in the art and are described in Winzeler et al. (1999) *Science* 285:901; Brenner (2000) *Genome Biol.* 1:1 Kumar et al. (2001) *Nature Rev.* 2:302; Giaever et al. (2004) *Proc. Natl. Acad. Sci. USA* 101:793; Eason et al. (2004) *Proc. Natl. Acad. Sci. USA* 101:11046; and Brenner (2004) *Genome Biol.* 5:240 each incorporated by reference in their entireties.

[089] As used herein, the term "primer" includes an oligonucleotide, either natural or synthetic, that is capable, upon forming a duplex with a polynucleotide template, of acting as a point of initiation of nucleic acid synthesis and being extended from its 3'

end along the template so that an extended duplex is formed. The sequence of nucleotides added during the extension process are determined by the sequence of the template polynucleotide. Usually primers are extended by a DNA polymerase. Primers usually have a length in the range of between 3 to 36 nucleotides, also 5 to 24 nucleotides, also from 14 to 36 nucleotides. Primers within the scope of the invention include orthogonal primers, amplification primers, constructions primers and the like. Pairs of primers can flank a sequence of interest or a set of sequences of interest. Primers and probes can be degenerate in sequence. Primers within the scope of the present invention bind adjacent to a target sequence. A "primer" may be considered a short polynucleotide, generally with a free 3' -OH group that binds to a target or template potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target. Primers of the instant invention are comprised of nucleotides ranging from 17 to 30 nucleotides. In one aspect, the primer is at least 17 nucleotides, or alternatively, at least 18 nucleotides, or alternatively, at least 19 nucleotides, or alternatively, at least 20 nucleotides, or alternatively, at least 21 nucleotides, or alternatively, at least 22 nucleotides, or alternatively, at least 23 nucleotides, or alternatively, at least 24 nucleotides, or alternatively, at least 25 nucleotides, or alternatively, at least 26 nucleotides, or alternatively, at least 27 nucleotides, or alternatively, at least 28 nucleotides, or alternatively, at least 29 nucleotides, or alternatively, at least 30 nucleotides, or alternatively at least 50 nucleotides, or alternatively at least 75 nucleotides or alternatively at least 100 nucleotides.

[090] According to one aspect, a number of sufficient barcode sequences is required to statistically ensure that each DNA or RNA in a sample has its own unique barcode. For example, a barcode having six bases would generate 4^6 or 4096 unique barcode sequences. According to one aspect, the number of RNA molecules is dependent upon cell type and one of skill in the art will reference readily available literature sources for estimates of the number of RNA molecules in a particular cell type. Accordingly, a statistically greater number of barcodes, or primer-barcode combinations or adapter-barcode combinations is required to facilitate the matching of a unique barcode for each RNA molecule. In embodiments where copy number

barcoding is applied to single cell gene expression profiling, a sufficient number or set of unique barcodes for unique labeling of a substantial fraction of RNA molecules in the sample is provided. Examples of single cell RNA samples include single *E. coli* cells which contain ~260,000 RNA molecules (Neidhardt et al., *Escherichia coli and Salmonella: Cell. Mol. Bio.*, Vol. 1, pp. 13-16, 1996), embryonic stem cells which contain ~10 pg mRNA/cell (~10^7 molecules) (Tang et al., *Nat. Protocols*, 5, 517, 2010), and neurons which contain ~50 pg mRNA/cell (5x10^7 molecules) (Uemura et al., *Experimental Neurobiology*, 65, 107, 1979).

[091] The barcoding process can be modeled analytically to calculate the barcode length required to uniquely barcode a substantial fraction of a sample of target nucleic acid molecules. A substantial fraction or percentage includes 80%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9% and higher. According to one aspect, a pool of random barcode sequences is provided. Consider a stock pool of barcodes with length L containing multiple copies of the 4^L barcodes. According to one embodiment, in order to barcode a sample of N target nucleic acid molecules, a sample containing M barcode molecules of random sequence is taken from a large barcode pool and added to the sample of N target nucleic acid molecules such that M > N. Because the identity of the M barcode molecules is random, the copy number distribution of the barcode sequences added to the sample is given by the Poisson Distribution where $P(k) = (M/4^L)^k e^{-M/4^{\wedge}L}/k!$ where k is the copy number. A random subset of these M barcode molecules will then label the N target nucleic acid molecules, a process which is also described by the Poisson Distribution. Hence, the distribution of target nucleic acid molecules for which a j-1 other target nucleic acid molecules share the same barcode sequence is given by $P(j) = (N/4^L)^j e^{-N/4^{\wedge}L}/j!$ Figure 8 shows a plot of the calculated fraction of molecules that are uniquely barcoded as a function of barcode length for four different sample sizes (N=10^5,10^6,10^7,10^8). Even for large samples containing 10^8 target nucleic acid molecules (similar to the case in which the RNA from a human cell is to be barcoded), a barcode length of L = 18 provides a unique barcode sequence for >99.9% of target molecules According to one embodiment, only target molecules with identical sequences need to be uniquely barcoded in order to count the number of target molecules present, as target molecules having a unique

sequence can be identified based on their unique nucleic acid sequence and do not necessarily need a unique barcode sequence.

[092] Barcodes within the scope of the present disclosure include nucleic acids of between about 3 and about 75 bases, about 5 and about 60 bases, about 6 and about 50 bases, about 10 and about 40 bases, or about 15 and about 30 bases. Barcodes can be comprised of any nucleic acid including DNA, RNA, and LNA. Random or pseudo-random barcode sequences can be synthesized using automated DNA synthesis technology that is known in the art (Horvath et al., *Methods Enzymol.*, 153, 314-326, 1987). Several commercially available services exist for the synthesis of randomly or pseudo-randomly barcoded nucleic acid oligonucleotides including Integrated DNA Technologies, Invitrogen, and TriLink Biotechnologies. In addition, there are enzymatic methods of synthesizing single-stranded nucleic acids. For example, non-template-directed enzymatic synthesis of nucleic acids can be carried out with poly(U) polymerase (for RNA) or terminal transferase (for DNA). For non-random barcode sets, methods of highly parallel DNA synthesis of specific, pre-determined DNA oligos are also known in the art such as the maskless array synthesis technology commercialized by Nimblegen. All of these methods are capable of adding a random, pseudo-random, or non-random barcode to a pre-determined sequence such as a primer or adapter.

[093] According to one aspect, a copy number barcode is 5'-conjugated to a primer such as a poly-T primer. The barcode of the present disclosure is referred to as a copy number barcode because the use of unique barcode sequences allows one of skill to determine the total number of nucleic acids within the sample and also the copy number of nucleic acids within a sample. A barcode sequence can be conjugated to a primer using methods known to those of skill in the art. Such methods include the combination of automated, random and directed nucleic acid synthesis where the random component generates a barcode and the directed component generates a specific primer (Horvath, S.J., Firca, J.R., Hunkapiller, T., Hunkapiller, M.W., Hood, L. An automated DNA synthesizer employing deoxynucleoside 3'-phosphoramidites. *Methods Enzymol.*, 153, 314-326, 1987), single-stranded ligation (Tessier, D.C.,

Brousseau, R., Vernet, T. Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase, *Anal. Biochem.*, 158, 171-178, 1986.), and double-stranded ligation (Meyer, M., Stenzel, U., Myles, S., Prufer, K., Hofreiter, M., Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucl. Acids Res.*, 35, e97, 2007).

[094] According to one aspect, a copy number barcode is conjugated to a hairpin primer. A variety of commercially available enzymes can be used to ligate DNA to RNA including CircLigase (Epicentre), CircLigase II (Epicentre), T4 RNA Ligase 1 (New England Biolabs), and T4 RNA Ligase 2 (New England Biolabs), allowing the ligation of specific DNA hairpin primers, as shown in Figure 2, to the 3' end of RNA for reverse transcription. In certain exemplary embodiments, DNA hairpin primers include a self-complementary region that folds into double-stranded DNA and serves as a DNA primer for reverse transcriptase. The DNA hairpin can also optionally include other features such as recognition sequences for enzymes that digest double-stranded DNA, copy number barcodes, sample barcodes, PCR adapters and the like. Because reverse transcriptase can replicate both DNA and RNA, the hairpin can include a single-stranded DNA 5' overhang with any additional sequence content.

[095] As shown in Figure 2-1, a hairpin primer includes a hairpin for priming reverse transcription on its 3' end along with a copy number barcode and a sample barcode on its 5' end. However, a hairpin primer could also include PCR adapters on the 5' end as shown in Figure 2-2. PCR adapters should be compatible with the library preparation protocols employed in an appropriate massively parallel sequencing platform. The inclusion of two PCR adapters eliminates downstream ligation steps that are otherwise typically needed for amplification and sequencing library preparation. Following RNA/DNA ligation and reverse transcription, the resultant cDNA can be circularized using, e.g., CircLigase (Epicentre), CircLigase II (Epicentre) or the like, allowing exponential amplification of the cDNA library via PCR. Alternatively, if circularization is undesirable, the hairpin primer can be designed to include only one of two PCR adapters. A subsequent ligation step can then be included to attach a second PCR adapter to the 3' end of cDNA.

[096] As shown in Figure 1, a linear pre-amplification step is provided where multiple copies of cDNA are made by repeated reverse transcription of the RNA with the primer-barcode conjugate. This method is useful when very low concentrations of the nucleic acids of interest are available (e.g. in single cell or diagnostic applications). Inefficiencies due to material loss can be mitigated without introducing bias by incorporating the linear pre-amplification step into the reverse transcription process.

[097] As shown in Figure 2-3, by including a recognition sequence for a nicking enzyme in the hairpin primer, multiple copies of cDNA can be generated by the inclusion of the corresponding nicking enzyme in the reverse transcriptase reaction mixture. Before reverse transcribing the ligated RNA into cDNA, reverse transcriptase will convert the single-stranded DNA recognition sequence into double-stranded DNA, allowing the nicking enzyme to generate a nick at the recognition site as shown in Figure 3. Reverse transcriptase can then use the nick as a priming site for additional replication because its strand-displacement activity allows the removal of the most recently generated cDNA copy from the RNA. Repeated cycles of nicking and reverse transcription results in linear amplification of RNA such that multiple copies of each barcoded cDNA are generated. The resultant cDNA library can be circularized and amplified using PCR. The same method of linear pre-amplification can be applied to a sample of uniquely barcoded DNA molecules. Instead of repeatedly reverse transcribing the barcoded DNA, repeated cycles of nicking and DNA replication by a strand-displacing DNA polymerase will result in linear amplification of DNA such that multiple copies of barcoded DNA are generated. Strand-displacing DNA polymerases are known in the art and include ϕ29 DNA polymerase, Klenow Fragment DNA polymerase, *Bst* Large Fragment DNA polymerase, Vent DNA polymerase, Deep Vent DNA polymerase, and 9°N DNA polymerase.

[098] Because an overwhelming amount of hairpin primer is included in certain of the ligation reactions described herein, it may be important to remove excess hairpin primer prior to PCR amplification. A circularization reaction performed prior to PCR results not only in circularized cDNA, but also in circularization of excess, linearly amplified 5' ends of the hairpin primer that were not ligated to RNA. The excess 5'

ends that are circularized could cause premature saturation of PCR because, although they are not attached to cDNA, they do include the two PCR adapters. The 5' ends of the hairpin primers can be designed to form a restriction or cut site following circularization to avoid exponential amplification during PCR and facilitate their removal by exonuclease-mediated digestion. For example, Figure 2-4 shows a hairpin primer containing not only a hairpin priming site for reverse transcriptase, a nicking enzyme recognition sequence, a copy number barcode, a sample barcode, and two PCR adapters, but also two separated halves of a restriction site. The first half of the restriction site occurs on the 3' end of the copy number barcode, and the second half occurs on the 5' end of the hairpin primer (i.e., the 5' end of the two PCR adapters). The single stranded 5' end of any excess hairpin primer that is not ligated to an RNA molecule will be linearly amplified by the combination of reverse transcriptase and a nicking enzyme. In addition, the resultant amplicons will be circularized in the same reaction that circularizes the cDNA library. However, for amplicons that do not include cDNA, the circularization reaction will join the two halves of the restriction site, allowing those amplicons to be selectively converted to linear DNA using a restriction enzyme. The restriction digest product will not be exponentially amplified by PCR. Furthermore, this linearized excess DNA can be eliminated by digestion with an exonuclease (e.g., Exonuclease I, which will digest linearized DNA but not circularized DNA).

[099] Primers for specific reverse transcription of an organism's RNA can be obtained from the organism's genomic DNA. Genomic DNA can be enzymatically fragmented and ligated to PCR adapters that include restriction sites for downstream isolation of primer-sized DNA fragments. The adapter-ligated fragments can be amplified by PCR using PCR primers that include copy-number barcodes. In certain aspects, this process results in a library of barcoded genomic fragments which can then be cut into smaller fragments by one or more restriction enzymes. The smaller fragments are referred to as a genomic primer pool which can serve as a set of specific primers for capturing and reverse transcribing the RNA of the organism from which the pool was isolated. Furthermore, the fragments comprising the genomic primer pool include not only a specific, genomic primer sequence, but also a copy number barcode.

[0100] Although random oligonucleotide sequences can serve as copy number barcodes, it may be desirable in some cases to place certain constraints on the set of sequences used for barcoding. For example, certain homopolymeric sequences and sequences with very high G+C or A+T content may be difficult to amplify and sequence. In addition, barcode sequences that include other specific sequences or the complements of specific sequences to be used in library preparation (e.g., PCR adapters or sample barcodes) may be undesirable in certain circumstances. For a given organism or sample type, design principles may be applied to generate an optimized set of copy number barcodes. An optimized set of barcodes would contain a minimal number of members capable of hybridizing efficiently to genetic material in the target organism or to adapters, amplification primers, capture primers, enzyme recognition sequences, or sequencing primers used in sample preparation, quality control steps, or sequencing. In addition, an optimized barcode set would not include barcode sequences containing large homopolymeric tracts (e.g. >5 identical bases in a row), G-quadraplexes, or highly GC- or AT-rich sequences (e.g. GC-fraction should lie between 30-70%). Because such an optimized barcode set cannot be generated by random DNA synthesis, methods of highly parallel, sequence-specific DNA synthesis can be used. Methods for highly parallel, maskless array synthesis such as those commercialized by Nimblegen, are known in the art.

[0101] In addition, computer based methods can be used to design an optimized barcode set *in silico* using software and design parameters as described herein. The optimized barcode set may then be synthesized using methods described herein and/or known to those of skill in the art.

[0102] As shown in Figure 1, cDNA with the unique barcode sequence is prepared for amplification using methods known to those of skill in the art. Such methods include, for example, circularization as described in WO/2010/094040, single-stranded adapter ligation as described in WO/2010/094040, hybridization of random primers as described in U.S. 6,124,120, or double-stranded adapter ligation as described in U.S. 7,741,463 each of which are hereby incorporated by reference in their entireties.

[0103] As shown in Figure 1, cDNA with the unique barcode sequences are amplified using methods known to those of skill in the art. In certain aspects, amplification is achieved using PCR. The term "polymerase chain reaction" ("PCR") of Mullis (U.S. Pat. Nos. 4,683,195, 4,683,202, and 4,965,188) refers to a method for increasing the concentration of a segment of a target sequence in a mixture of nucleic acid sequences without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the nucleic acid sequence mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a polymerase (e.g., DNA polymerase). The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle;" there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

[0104] With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR

amplifications. Methods and kits for performing PCR are well known in the art. PCR is a reaction in which replicate copies are made of a target polynucleotide using a pair of primers or a set of primers consisting of an upstream and a downstream primer, and a catalyst of polymerization, such as a DNA polymerase, and typically a thermally-stable polymerase enzyme. Methods for PCR are well known in the art, and taught, for example in MacPherson et al. (1991) PCR 1: A Practical Approach (IRL Press at Oxford University Press). All processes of producing replicate copies of a polynucleotide, such as PCR or gene cloning, are collectively referred to herein as replication. A primer can also be used as a probe in hybridization reactions, such as Southern or Northern blot analyses.

[0105] The expression "amplification" or "amplifying" refers to a process by which extra or multiple copies of a particular polynucleotide are formed. Amplification includes methods such as PCR, ligation amplification (or ligase chain reaction, LCR) and amplification methods. These methods are known and widely practiced in the art. See, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202 and Innis et al., "PCR protocols: a guide to method and applications" Academic Press, Incorporated (1990) (for PCR); and Wu et al. (1989) Genomics 4:560-569 (for LCR). In general, the PCR procedure describes a method of gene amplification which is comprised of (i) sequence-specific hybridization of primers to specific genes within a DNA sample (or library), (ii) subsequent amplification involving multiple rounds of annealing, elongation, and denaturation using a DNA polymerase, and (iii) screening the PCR products for a band of the correct size. The primers used are oligonucleotides of sufficient length and appropriate sequence to provide initiation of polymerization, i.e. each primer is specifically designed to be complementary to each strand of the genomic locus to be amplified.

[0106] Reagents and hardware for conducting amplification reaction are commercially available. Primers useful to amplify sequences from a particular gene region are preferably complementary to, and hybridize specifically to sequences in the target region or in its flanking regions and can he prepared using the polynucleotide

sequences provided herein. Nucleic acid sequences generated by amplification can be sequenced directly.

[0107] When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, the reaction is called "annealing" and those polynucleotides are described as "complementary". A double-stranded polynucleotide can be complementary or homologous to another polynucleotide, if hybridization can occur between one of the strands of the first polynucleotide and the second. Complementarity or homology (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonding with each other, according to generally accepted base-pairing rules.

[0108] The terms "reverse-transcriptase PCR" and "RT-PCR" refer to a type of PCR where the starting material is mRNA. The starting mRNA is enzymatically converted to complementary DNA or "cDNA" using a reverse transcriptase enzyme. The cDNA is then used as a template for a PCR reaction.

[0109] The terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

[0110] The term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.). Amplification methods include PCR methods known to those of skill in the art and also include rolling circle amplification (Blanco et al., *J. Biol. Chem.*, 264, 8935-8940, 1989), hyperbranched rolling circle amplification (Lizard et al., *Nat. Genetics*, 19, 225-232, 1998), and loop-mediated isothermal amplification

(Notomi et al., *Nuc. Acids Res.*, 28, e63, 2000) each of which are hereby incorporated by reference in their entireties.

[0111] The cDNA samples can be amplified by polymerase chain reaction (PCR) including emulsion PCR and single primer PCR in the methods described herein. For example, the cDNA samples can be amplified by single primer PCR. The CDS can comprise a 5' amplification primer sequence (APS), which subsequently allows the first strand of cDNA to be amplified by PCR using a primer that is complementary to the 5' APS. The TSO can also comprise a 5' APS, which can be at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, or 70%, 80%, 90% or 100% identical to the 5' APS in the CDS. This means that the pooled cDNA samples can be amplified by PCR using a single primer (i.e. by single primer PCR), which exploits the PCR suppression effect to reduce the amplification of short contaminating amplicons and primer-dimers (Dai et al., *J Biotechnol* 128(3):435-43 (2007)). As the two ends of each amplicon are complementary, short amplicons will form stable hairpins, which are poor templates for PCR. This reduces the amount of truncated cDNA and improves the yield of longer cDNA molecules. The 5' APS can be designed to facilitate downstream processing of the cDNA library. For example, if the cDNA library is to be analyzed by a particular sequencing method, e.g. Applied Biosystems' SOLiD sequencing technology, or Illumina's Genome Analyzer, the 5' APS can be designed to be identical to the primers used in these sequencing methods. For example, the 5' APS can be identical to the SOLiD Pl primer, and/or a SOLiD P2 sequence inserted in the CDS, so that the PI and P2 sequences required for SOLiD sequencing are integral to the amplified library.

[0112] For emulsion PCR, an emulsion PCR reaction is created by vigorously shaking or stirring a "water in oil" mix to generate millions of micron-sized aqueous compartments. The DNA library is mixed in a limiting dilution either with the beads prior to emulsification or directly into the emulsion mix. The combination of compartment size and limiting dilution of beads and target molecules is used to generate compartments containing, on average, just one DNA molecule and bead (at the optimal dilution many compartments will have beads without any target) To

facilitate amplification efficiency, both an upstream (low concentration, matches primer sequence on bead) and downstream PCR primers (high concentration) are included in the reaction mix. Depending on the size of the aqueous compartments generated during the emulsification step, up to $3x10^9$ individual PCR reactions per µl can be conducted simultaneously in the same tube. Essentially each little compartment in the emulsion forms a micro PCR reactor. The average size of a compartment in an emulsion ranges from sub- micron in diameter to over a 100 microns, depending on the emulsification conditions.

[0113] "Identity," "homology" or "similarity" are used interchangeably and refer to the sequence similarity between two nucleic acid molecules. Identity can be determined by comparing a position in each sequence which can be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of identity between sequences is a function of the number of matching or identical positions shared by the sequences. An unrelated or non-homologous sequence shares less than 40% identity, or alternatively less than 25% identity, with one of the sequences of the present invention.

[0114] A polynucleotide has a certain percentage (for example, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99%) of "sequence identity" to another sequence means that, when aligned, that percentage of bases are the same in comparing the two sequences. This alignment and the percent sequence identity or homology can be determined using software programs known in the art, for example those described in Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., (1993). Preferably, default parameters are used for alignment. One alignment program is BLAST, using default parameters. In particular, programs are BLASTN and BLASTP, using the following default parameters: Genetic code = standard; filter = none; strand = both; cutoff- 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + SwissProtein + SPupdate + PIR.

Details of these programs can be found at the National Center for Biotechnology Information.

[0115] As shown in Figure 1, the amplified cDNA is sequenced and analyzed using methods known to those of skill in the art. In certain exemplary embodiments, RNA expression profiles are determined using any sequencing methods known in the art. Determination of the sequence of a nucleic acid sequence of interest can be performed using a variety of sequencing methods known in the art including, but not limited to, sequencing by hybridization (SBH), sequencing by ligation (SBL) (Shendure et al. (2005) *Science* 309:1728), quantitative incremental fluorescent nucleotide addition sequencing (QIFNAS), stepwise ligation and cleavage, fluorescence resonance energy transfer (FRET), molecular beacons, TaqMan reporter probe digestion, pyrosequencing, fluorescent *in situ* sequencing (FISSEQ), FISSEQ beads (U.S. Pat. No. 7,425,431), wobble sequencing (PCT/US05/27695), multiplex sequencing (U.S. Serial No. 12/027,039, filed February 6, 2008; Porreca et al (2007) *Nat. Methods* 4:931), polymerized colony (POLONY) sequencing (U.S. Patent Nos. 6,432,360, 6,485,944 and 6,511,803, and PCT/US05/06425); nanogrid rolling circle sequencing (ROLONY) (U.S. Serial No. 12/120,541, filed May 14, 2008), allele-specific oligo ligation assays (e.g., oligo ligation assay (OLA), single template molecule OLA using a ligated linear probe and a rolling circle amplification (RCA) readout, ligated padlock probes, and/or single template molecule OLA using a ligated circular padlock probe and a rolling circle amplification (RCA) readout) and the like. High-throughput sequencing methods, e.g., using platforms such as Roche 454, Illumina Solexa, ABI-SOLiD, Helicos, Complete Genomics, Polonator platforms and the like, can also be utilized. A variety of light-based sequencing technologies are known in the art (Landegren et al. (1998) *Genome Res.* 8:769-76; Kwok (2000) *Pharmacogenomics* 1:95-100; and Shi (2001) *Clin. Chem.* 47:164-172).

[0116] The method of preparing a cDNA library described herein can further comprise processing the cDNA library to obtain a library suitable for sequencing. As used herein, a library is suitable for sequencing when the complexity, size, purity or the like of a cDNA library is suitable for the desired screening method. In particular, the

cDNA library can be processed to make the sample suitable for any high-throughout screening methods, such as Applied Biosystems' SOLiD sequencing technology, or Illumina's Genome Analyzer. As such, the cDNA library can be processed by fragmenting the cDNA library (e.g. with DNase) to obtain a short- fragment 5 '-end library. Adapters can be added to the cDNA, e.g. at one or both ends to facilitate sequencing of the library. The cDNA library can be further amplified, e.g. by PCR, to obtain a sufficient quantity of cDNA for sequencing.

[0117] Embodiments of the invention provide a cDNA library produced by any of the methods described herein. This cDNA library can be sequenced to provide an analysis of gene expression in single cells or in a plurality of single cells.

[0118] Embodiments of the invention also provide a method for analyzing gene expression in a plurality of single cells, the method comprising the steps of preparing a cDNA library using the method described herein and sequencing the cDNA library. A "gene" refers to a polynucleotide containing at least one open reading frame (ORF) that is capable of encoding a particular polypeptide or protein after being transcribed and translated. Any of the polynucleotide sequences described herein can be used to identify larger fragments or full- length coding sequences of the gene with which they are associated. Methods of isolating larger fragment sequences are known to those of skill in the art.

[0119] As used herein, "expression" refers to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA is subsequently being translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression can include splicing of the mRNA in an eukaryotic cell.

[0120] The cDNA library can be sequenced by any suitable screening method. In particular, the cDNA library can be sequenced using a high-throughout screening method, such as Applied Biosystems' SOLiD sequencing technology, or Illumina's Genome Analyzer. In one aspect of the invention, the cDNA library can be shotgun sequenced.

The number of reads can be at least 10,000, at least 1 million, at least 10 million, at least 100 million, or at least 1000 million. In another aspect, the number of reads can be from 10,000 to 100,000, or alternatively from 100,000 to 1 million, or alternatively from 1 million to 10 million, or alternatively from 10 million to 100 million, or alternatively from 100 million to 1000 million. A "read" is a length of continuous nucleic acid sequence obtained by a sequencing reaction.

[0121] "Shotgun sequencing" refers to a method used to sequence very large amount of DNA (such as the entire genome). In this method, the DNA to be sequenced is first shredded into smaller fragments which can be sequenced individually. The sequences of these fragments are then reassembled into their original order based on their overlapping sequences, thus yielding a complete sequence. "Shredding" of the DNA can be done using a number of difference techniques including restriction enzyme digestion or mechanical shearing. Overlapping sequences are typically aligned by a computer suitably programmed. Methods and programs for shotgun sequencing a cDNA library are well known in the art.

[0122] The expression profiles described herein are useful in the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining the expression profile of nucleic acid sequences (e.g., RNAs), in order to determine whether an individual is at risk of developing a disorder and/or disease. Such assays can be used for prognostic or predictive purposes to thereby prophylactically treat an individual prior to the onset of the disorder and/or disease. Accordingly, in certain exemplary embodiments, methods of diagnosing and/or prognosing one or more diseases and/or disorders using one or more of expression profiling methods described herein are provided.

[0123] Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds administered either to inhibit or to treat or prevent a disorder and/or disease) on the expression profile of nucleic acid sequences (e.g.,

RNAs) in clinical trials. Accordingly, in certain exemplary embodiments, methods of monitoring one or more diseases and/or disorders before, during and/or subsequent to treatment with one or more agents using one or more of expression profiling methods described herein are provided.

[0124] Monitoring the influence of agents (e.g., drug compounds) on the level of expression of a marker of the invention can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent to affect an expression profile can be monitored in clinical trials of subjects receiving treatment for a disease and/or disorder associated with the expression profile. In certain exemplary embodiments, the methods for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting one or more expression profiled in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting one or more expression profiles in the post-administration samples; (v) comparing the one or more expression profiled in the pre-administration sample with the one or more expression profiles in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly.

[0125] As used herein, the term "biological sample" is intended to include, but is not limited to, tissues, cells, biological fluids and isolates thereof, isolated from a subject, as well as tissues, cells and fluids present within a subject. Many expression detection methods use isolated RNA. Any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from biological samples (see, e.g., Ausubel et al., ed., Current Protocols in Molecular Biology, John Wiley & Sons, New York 1987-1999). Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (1989, U.S. Pat. No. 4,843,155).

[0126] The expression profiling methods described herein allow the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a variety of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression per se and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[0127] In another embodiment, the time course of expression of one or more nucleic acid sequences (e.g., genes, mRNAs and the like) in an expression profile can be monitored. This can occur in various biological contexts, as disclosed herein, for example development of a disease and/or disorder, progression of a disease and/or disorder, and processes, such a cellular alterations associated with the disease and/or disorder.

[0128] The expression profiling methods described herein are also useful for ascertaining the effect of the expression of one or more nucleic acid sequences (e.g., genes, mRNAs and the like) on the expression of other nucleic acid sequences (e.g., genes, mRNAs and the like) in the same cell or in different cells. This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[0129] The expression profiling methods described herein are also useful for ascertaining differential expression patterns of one or more nucleic acid sequences (e.g., genes, mRNAs and the like) in normal and abnormal cells. This provides a battery of nucleic acid sequences (e.g., genes, mRNAs and the like) that could serve as a molecular target for diagnosis or therapeutic intervention.

[0130] In certain exemplary embodiments, electronic apparatus readable media comprising one or more expression profiles described herein is provided. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy disks, hard disk storage medium, and magnetic tape; optical storage media such as compact disc; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon one or more expression profiles described herein.

[0131] As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatuses suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

[0132] As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising one or more expression profiles described herein.

[0133] A variety of software programs and formats can be used to store the marker information of the present invention on the electronic apparatus readable medium.

For example, the marker nucleic acid sequence can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of data processor structuring formats (e.g., text file or database) may be employed in order to obtain or create a medium having recorded thereon one or more expression profiles described herein.

[0134] By providing one or more expression profiles described herein in readable form, one can routinely access the expression profile information for a variety of purposes. For example, one skilled in the art can use the one or more expression profiles described herein in readable form to compare a target expression profile with the one or more expression profiles stored within the data storage means. Search means are used to identify similarities and/or differences between two or more expression profiles.

[0135] **Figure 10(A)** depicts the general concept of digital counting by random labeling of all target nucleic acid molecules in a sample with unique barcode sequences. Assume the original sample contains two cDNA sequences, one with three copies and another with two copies. An overwhelming number of unique barcode sequences are added to the sample in excess, and five are randomly ligated to the cDNA molecules. Ideally, each cDNA molecule in the sample receives a unique barcode sequence. After removing the excess barcodes, the barcoded cDNA molecules are amplified by PCR. Because of intrinsic noise and sequence-dependent bias, the barcoded cDNA molecules may be amplified unevenly. Consequently, after the amplicons are sequenced, it may appear that there are three copies of cDNA1 for every four copies of cDNA2 based on the relative number of reads for each sequence. However, the ratio in the original sample was 3:2, which is accurately reflected in the relative number of unique barcodes associated with each cDNA sequence. In the implementation of the method depicted in Figure 10(A), it may be advantageous to randomly ligate both ends of each phosphorylated cDNA fragment to a barcoded phosphorylated Illumina Y-shaped adapter as shown in Figure 10(B). Note that the single T and A overhangs present on the barcodes and cDNA, respectively, are to enhance ligation efficiency. After this

step, the sample is amplified by PCR and prepared for sequencing using the standard Illumina library protocol. For each amplicon, both barcode sequences and both strands of the cDNA sequence are read using paired-end deep sequencing. According to one aspect, the paired-end strategy, i.e. attaching a barcode to each end of the nucleic acid in a sample, reduces the number of barcodes that must be designed and synthesized while allowing conventional paired-end library protocols and providing long-range sequence information that improves mapping accuracy.

[0136] It is to be understood that the embodiments of the present invention which have been described are merely illustrative of some of the applications of the principles of the present invention. Numerous modifications may be made by those skilled in the art based upon the teachings presented herein without departing from the true spirit and scope of the invention. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference in their entirety for all purposes.

[0137] The following examples are set forth as being representative of the present invention. These examples are not to be construed as limiting the scope of the invention as these and other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

EXAMPLE I

Single Cell RNA Quantification with Copy Number Barcodes

Cell Selection by Laser Dissection Microscopy

[0138] Prior to cell selection, the intracellular RNA is stabilized by adding 0.5 mL of RNALater (Ambion) to every 0.1 mL of cell culture. A cell is then selected, cut from a culture dish, and dispensed in a tube using a laser dissection microscope (LMD-6500, Leica). The cells are plated onto a membrane-coated culture dish and observed using bright field microscopy with a 10x objective (Leica). A UV laser is then used to cut the membrane around an individually selected cell such that it falls into the cap of a PCR tube containing 20 μL of buffer. The captured cell is then thermally lysed.

Genomic DNA Removal

[0139] Genomic DNA can be removed from the sample by the addition of 1 μL of 0.1 U/μL DNase I (New England BioLabs) and incubation at 37 C for 15 minutes. The reaction is the quenched by the addition of EDTA to a final concentration of 5 mM and heat inactivation at 75 C for 10 minutes.

Reverse Transcription and Copy Number Barcoding using RNA Ligation with Linear Amplification

[0140] To facilitate reverse transcription, a DNA oligonucleotide with the following sequence is ligated to each RNA molecule in the sample:

[0141] 5'-P
GCAGATCGGAAGAGGCTCGTTGAGGGAACAGGTTCAGAGTTCTACAGT
CCGACGATCGGCNNNNNNNNNNNNNNNNNNNTCCAACCGAAGAGCGG
TGCACCGTCCGAGTCGACGTCTGGATCCTGTTCTTTCTCAGGATCCAGAC
GTCGACTCGGACGGTGCACCGCTCTTCG-3' (SEQ ID NO:1)

[0142] This oligonucleotide includes a hairpin primer for reverse transcription, two PCR adapters, the recognition sequence of a nicking enzyme, a sample barcode (for massively parallel sequencing), and a 20-base copy number barcode, and the 5' end is phosphorylated. The hairpin primer set is added to the single cell RNA sample to a final concentration of 1 μM such that nearly every RNA molecule in the sample will be conjugated to a different copy number barcode. The hairpin primers are ligated to the RNA molecules using T4 RNA Ligase 1 under the following conditions:

1x T4 RNA Ligase 1 Reaction Buffer (New England Biolabs, Ipswich, MA)
0.7 units/μL T4 RNA Ligase 1
10% (v/v) DMSO

[0143] The reaction mixture is incubated for 12 hours at 16 °C. Lambda Exonuclease and Exonuclease I are then added to the sample at final concentrations of 1 unit/μL, and the sample is incubated at 37 °C for one hour to digest excess hairpin primer. The two exonucleases are then heat inactivated by incubation at 80 °C for 20 minutes.

[0144] The hairpin primed RNA templates are then linearly amplified by the combination of a reverse transcriptase with strand-displacement activity and a nicking enzyme. The nicking enzyme cuts at a recognition sequence adjacent to the hairpin primer whenever reverse transcriptase generates double-stranded DNA at the recognition site. This cut site allows reverse transcriptase to copy the RNA template repeatedly, removing the previously generated cDNA with its strand-displacement activity. To accomplish this, MonsterScript (Epicentre) and Nt.BspQI (New England Biolabs) are added to the sample at final concentrations of 0.5 units/μL and 2 units/μL, respectively, along with 0.2 mM dNTPs. The reaction mixture is then incubated for one hour at 50 °C.

[0145] The linearly amplified cDNA is then circularized using CircLigase II (Epicentre Biotechnologies, Madison, WI). Manganous chloride, betaine, and CircLigase II are added to the sample at final concentrations of 2.5 mM, 1 M, and 5 units/μL, respectively. The new reaction mixture is then incubated for one hour at 60 °C.

Sequencing Library Preparation and Massively Parallel Sequencing

[0146] Using the two adapters included in each cDNA molecule, the cDNA is amplified by PCR. This is accomplished using the manufacturer's instructions for library preparation and varies slightly for different massively parallel sequencing platforms (e.g., Roche/454 Life Sciences, SOLiD by Life Technologies, HiSeq 2000 or Genome Analyzer by Illumina). In this particular example, Illumina adapter sequences are included in the cDNA library such that the circularized templates can be amplified using the standard Illumina library preparation kit (e.g. TruSeq) and sequenced on a HiSeq 2000.

EXAMPLE II

DNA Quantification Using Copy Number Barcodes

[0147] Eight DNA template oligos (135 bases) were purchased from Integrated DNA Technologies along with two primers:

[0148] Template A: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNNNN NNNNNNNNTGGTGGAGCTGGCGGGAGTTGAACCCGCGTCCGAAATTCCT ACATCCTCGGTACTACATGGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:2)

[0149] Template B: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNN NNNNNNNNNNTCGGGCCGGGGGTTGGGCCAGGCTCTGAGGTGTGGGGGA TTCCCCCATGCCCCCCGCCGTGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:3)

[0150] Template C: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNN NNNNNNNNNNTTATAAATACCGGCCCCGGCGGAAAACCAAGACGCTCAT GAAGAAGGATAAGTACACGCTGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:4)

[0151] Template D: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNN NNNNNNNNNNCGCCGCGGGGTGCACCGTCCGGACCCTGTTTTCAGGGTC CGGACGGTGCACCCCGCGGCGGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:5)

[0152] Template E: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNN NNNNNNNNNNCAAGCAGAAGACGGCTCCGGGACCGTCCGGACCCTGTTT TCAGGGTCCGGACGGTCCCGGGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:6)

[0153] Template F: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNN NNNNNNNNNNGTTGCAGAAGACGGCTCCGGGACCGTCCGGACCCTGTTT TCAGGGTCCGGACGGTCCCGGGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:7)

[0154] Template G: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNN NNNNNNNNNNNCGCCGCGGTGCACCTTTTGGTGCACCGCGGCGCCCGCG

TCCGAAATTCCTACATCCTCGGGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:8)

[0155] Template H: 5'-ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNN NNNNNNNNNNNGTGAGAGAGTGAGCGAGACAGAAAGAGAGAGAAGTGC ACCAGCGAGCCGGGGCAGGAAGAGCCGTCGTATGCCGTCTTCTGCTTG-3' (SEQ ID NO:9)

[0156] Primer 1:

5'-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA-3' (SEQ ID NO:10)

[0157] Primer 2:

5'-CAAGCAGAAGACGGCATACGACGGC-3' (SEQ ID NO:11)

[0158] "N" designates a random base. The 3' and 5' ends of each template are complementary to Primer 1 and the complement of Primer 2, respectively. Each of the eight DNA oligos includes 16 random bases which serve as a copy number barcode. The templates are diluted identically in two different tubes containing a PCR Master Mix such that the average copy number of each template is as follows:

Templates A-D: 1 copy per tube
Template E: 10 copies per tube
Template F: 100 copies per tube
Template G: 10,000 copies per tube
Template H: 1,000,000 copies per tube.

[0159] The PCR Master Mix consists of:

0.5 μM Primer 1
0.5 μM Primer 2
0.2 mM dNTPs (New England Biolabs, Ipswich, MA)

1x Phusion HF Buffer (New England Biolabs)

0.02 units/μL Phusion DNA Polymerase (New England Biolabs)

at a final volume of 50 μL.

[0160] The two PCR samples are then thermocycled as follows:

1) 98 C for 30 s
2) 98 C for 10 s
3) 60 C for 30 s
4) 72 C for 30 s
5) Repeat steps 2-4 19 times.
6) 72 C for 10 minutes.

[0161] The two PCR samples are combined with 20 μL of ExoSAP-IT PCR Product Cleanup mixture (USB) and incubated for 15 minutes at 37 C followed by heat inactivation for 15 minutes at 80 C. Each of the two samples is then sequenced in one lane of a HiSeq 2000 (Illumina) sequencer according to the manufacturer's instructions resulting in approximately 10^8 reads per sample. For each template sequence, the number of different copy number barcode sequences is counted to determine the copy number of the template sequence in the original sample.

EXAMPLE III

Generation of a Genomic Primer Pool with Copy Number Barcodes

[0162] Genomic DNA is fragmented using Fragmentase (New England Biolabs) by combining 5 μg of purified genomic DNA (Figure 4B), 10 μL of NEBNext dsDNA Fragmentase, 1x NEBNext dsDNA Fragmentation Reaction Buffer (New England Biolabs), and 0.1 mg/mL BSA in a final volume of 100 μL. The reaction mixture is then incubated at 37 C for 30 minutes, quenched by the addition of 100 mM EDTA, purified on a DNA purification column (Zymo Research), and eluted in a final volume of 35 μL. Figure 4B depicts an electrophoretic gel in which samples taken from different time points in the Fragmentase reaction were run in different lanes of the gel. The first and second lanes contain standard DNA ladder sequences, and the third, fourth, and fifth lanes contain samples taken 20 minutes, 30 minutes, and 40 minutes from the initiation of DNA fragmentation, respectively. The gel shows that the average DNA length in the sample decreases as the reaction progresses.

[0163] Fragment end repair is accomplished using the NEBNext End Repair Enzyme Mix from New England Biolabs. The entire 35 μL of purified DNA fragments from the previous step are combined with 5 μL of NEBNext End Repair Enzyme Mix, 1x NEBNext End Repair Reaction Buffer, and *E. coli* Ligase (New England Biolabs) at 0.1 units/μL in a final volume of 100 μL. The reaction mixture is then incubated for 30 minutes at 20 C followed by purification with a DNA purification column (Zymo Research) and elution into a final volume of 42 μL. Deoxyadenosine tails (dA-tails) are then added to the purified, end-repaired DNA by combining all 42 μL of purified DNA with 1x NEBNext dA-Tailing Reaction Buffer and 3 μL Klenow fragment exo- (New England Biolabs) in a final volume of 50 μL. This reaction mixture is then incubated for 30 minutes at 37 C followed by purification with a DNA purification column (Zymo Research) and elution into a final volume of 8 μL.

[0164] Figure 4A shows a schematic of a DNA fragment after adapter ligation. Two PCR adapter oligonucleotides are ligated to the 5'-end and 3'-end of each DNA fragment in the sample. As shown in Figure 4A, the two PCR adapters are not only complementary to a set of PCR primers, they also contain a recognition sequence for MmeI, a restriction enzyme whose cut site is ~20 bases away from the recognition sequence. Oligonucleotides that include PCR adapters at their 5' end and a recognition site for the restriction enzyme MmeI at their 3' end are ligated onto the purified, dA-tailed genomic DNA fragments using Quick Ligase (New England Biolabs). This is accomplished by combining all 8 μL of purified DNA from the previous step with the oligonucleotides at a final concentration of 2 μM, 5 μL of Quick T4 DNA Ligase, and 1x Quick Ligation Reaction Buffer in a final volume of 50 μL. This reaction mixture is then incubated for 15 minutes at 20 C and loaded onto a 1.5% agarose gel which is run at 120 V for 50 minutes. The gel is stained with SybrSafe (Invitrogen), and the band that appears on the gel between 300 and 400 bp is then cut and the DNA is isolated from the gel using a gel purification kit (Qiagen) according to the manufacturer's instructions.

[0165] Following adapter ligation, the genomic fragments are amplified with PCR. The PCR primers used in this amplification step include copy number barcodes on their

5' ends so that the final amplicons are randomly or pseudo-randomly barcoded. In addition, the PCR primers can be biotinylated to facilitate isolation of the genomic primer pool in the final step of this protocol. For PCR, the purified, adapter-ligated genomic fragments are diluted to a final concentration of 100 fM and combined with 0.5 μL of Phusion DNA polymerase (New England Biolabs), 1x Phusion High Fidelity Buffer (New England Biolabs), 0.2 mM dNTPs, and 2 μM primers in a final volume of 50 μL. The PCR mixture is then thermocycled as follows:

1) 98 C for 30 s
2) 98 C for 5 s
3) 68 C for 15 s
4) 72 C for 10 s
5) Repeat steps 2-4 24 times.
6) 72 C for 10 minutes.

[0166] The resulting PCR amplicons are then purified with a DNA purification column (Zymo Research).

[0167] Figure 4C depicts an electrophoretic gel with lanes corresponding to samples in which different restriction enzymes were used to generate genomic primers. The first, second, and third lanes contain digestion products for BpuE1, BpmI, and MmeI, respectively. In each case, the lowest band corresponds to the digestion product which can be used as a pool of genomic primers. In order to isolate the genomic primer pool, the PCR amplicons are cut using MmeI, a restriction enzyme that cuts the double-stranded amplicons 20 bases from its recognition site, as shown in Figure 4C. The recognition site for MmeI was included in the oligonucleotides that were ligated to the genomic fragments such that digestion by MmeI will result in fragments that include a copy number barcode, a PCR adapter, a single-A base, and 19 bases of the genomic fragment. The restriction digest mixture includes 9 μL of purified PCR product, 1x NEB Buffer 4, 0.2 units/μL MmeI (New England Biolabs), and 50 μM S-adenosylmethionine in a final volume of 20 μL. The reaction mixture is incubated at 37 C for one hour followed by purification of single stranded fragments using streptavidin-coated magnetic beads according to the manufacturer's instructions for Roche/454 library preparation.

EXAMPLE IV

DNA Counting with a Copy Number Barcode using Sanger Sequencing

[0168] A series of DNA templates that either contained or did not contain a copy number barcode were amplified in a single tube by PCR with a common set of primers. This was accomplished by combining 1x Taq Master Mix (New England Biolabs), 0.5 μM of each PCR primer, and 10 fM of each DNA template in a final volume of 20 μL. The reaction mixture was thermocycled under the following conditions:

1) 94 C for 10 min
2) 94 C for 30 s
3) 58 C for 30 s
4) 72 C for 4 s
5) Repeat steps 2-4 29 times.
6) 72 C for 10 minutes.

[0169] The two DNA templates and PCR primers had the following sequences:

[0170] Template 1 (with copy number barcode):

5'-
CCCTACACGACGCTCTTCCGATCTNNNNNNAATGATACGGCGACCACCGA
G ATCTACACT-3' (SEQ ID NO:12)

[0171] Template 2 (without copy number barcode):

5'-CCCTACACGACGCTCTTCCGATCTAGCTCAAATGATACGGCGACCACCG
AGATCTACACT-3' (SEQ ID NO:13)

[0172] PCR Primer 1: 5'-AGTGTAGATCTCGGTGGTCGCCG-3' (SEQ ID NO:14)

[0173] PCR Primer 2: 5'-CCCTACACGACGCTCTTCCGATC-3' (SEQ ID NO:15)

[0174] The amplicons were then cloned using a TA cloning kit (Invitrogen) and sent out for Sanger sequencing (Genewiz). As shown in Figure 5, variant sequenced at the copy

number barcode site were only found when the DNA template and copy number barcode were used, showing that there are at least seven DNA copies in the sample.

EXAMPLE V

Copy Number Barcode Incorporation and RNA Counting

[0175] Reverse transcription was performed using DNA primers that either contained or did not contain a copy number barcode just as in the two cases shown in Figure 6. In this example, there are two copies of RNA in the original solution. As a first step, reverse transcription is performed by using DNA primer without (above) and with (bottom) CNB (6 bases in Figure 6). The CNB is actually a random nucleotide sequence which can be sequenced. The cDNA generated by reverse transcription is amplified and sequenced. By counting the CNBs, one can learn how many copies of RNA were in the original solution. This technique can be applied to a mixture of multiple RNA sequences in a single tube.

[0176] RNA was first obtained using an *in vitro* transcription kit (Ambion) to extract RNA from the ϕX174 viral genome with the following sequence:

5'-
AATCGCGTAGAGGCTTTGCTATTCAGCGTTTGATGAATGCAATGCGACAG
GCTCATGCTGATGGTTGGTTTATCGTTTTTGACACTCTCACGTTGGCTGAC
GACCGATTAGAGGCGTTTTATGATAATCCCAATGCTTTGCGTGACTATTTT
CGTGATATTGGTCGTATGGTTCTTGCTG-3' (SEQ ID NO:25)

[0177] The reverse transcription primer that includes a copy number barcode has the following sequence:

5'-
GAATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC
T TCCGATCTNNNNNNCGGTCGTCAGCCAACGTGAGAGTG-3' (SEQ ID NO:26)

[0178] This primer is phosphorylated on its 5' end. The sequence of the reverse transcription primer that does not include a copy number barcode is as follows:

[0179] 5'-GAATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCTAGCTGTCGGTCGTCAGCCAACGTGAGAGTG-3' (SEQ ID NO:27)

[0180] The primers at a final concentration of 2.2 μM were first annealed to the RNA sample at a final concentration of 2.1 μM in a total volume of 4.5 μL by heating the sample to 90 C and cooling gradually to 25 C, pausing every 5 C for one minute. The RNA sample was then reverse transcribed by combining 1x First-Strand Buffer (Invitrogen), 5 mM DTT, 0.5 mM dNTPs, and SuperScript III (Invitrogen, Carlsbad, CA) at a final concentration of 20 units/μL. The reaction mixture was then incubated at 37 C for 30 minutes. The sample was then treated with Exonuclease I (New England Biolabs) at a final concentration of 2 units/μL and incubated for 60 minutes at 37 C. The exonuclease was then inactivated by adding EDTA to a final concentration of 7.7 mM and incubating at 80 C for 25 minutes. The RNA was then digested by combining RNase H (Invitrogen) at 0.06 units/μL, RNase A (Qiagen) at 0.28 mg/mL, and 5.6 mM magnesium chloride. The reaction mixture was incubated for 30 minutes at 37 C followed by purification with a PuriZymo column (Zymo Research Corporation, Irvine, CA) and elution into a final volume of 6 μL. The final concentration of cDNA was found to be 1.6 μM.

[0181] The purified cDNA was circularized by combining 0.4 μM cDNA with 2.5 mM manganous chloride, 1 M betaine, 1x CIRCLIGASE™ II Reaction Buffer (Epicentre), and 0.1 μM CIRCLIGASE™ II (Epicentre). The reaction mixture was incubated for 1.5 hours at 60 C followed by heat inactivation for 10 minutes at 80 C. Any remaining linear DNA was then removed by the addition of Exonuclease I at a final concentration of 0.5 units/μL and incubation at 37 C for 30 minutes followed by heat inactivation for 20 minutes at 80 C.

[0182] The cDNAs were then amplified by PCR using two primers:

[0183] 5'-AGTGTAGATCTCGGTGGTCGCCG-3' (SEQ ID NO:28)

[0184] 5'-CCCTACACGACGCTCTTCCGATC-3' (SEQ ID NO:29)

[0185] By combining the cDNA with primers at 0.5 μM, dNTPs at 0.2 mM, 1x Phusion High Fidelity Buffer (New England Biolabs), and 0.5 μL of Phusion DNA polymerase (New England Biolabs) in a final volume of 20 μL. The reaction mixture was thermocycled under the following conditions:

1) 98 C for 30 s
2) 98 C for 10 s
3) 58 C for 15 s
4) 72 C for 35 s
5) Repeat steps 2-4 29 times.
6) 72 C for 8 minutes.

[0186] The PCR product was gel purified on a 1.5% agarose gel run at 120 V for 30 minutes followed by purification with a DNA purification column (Zymo Research Corporation). The purified PCR product was then cloned using a TA Cloning Kit (Invitrogen) and sent out for Sanger sequencing (Genewiz, South Plainfield, NJ). Just as in the case of DNA counting, variant sequences at the copy number barcode site were found only when the DNA primer containing the copy number barcode was used, showing that there are at least eight different RNA copies in the original sample. See Figure 7.

EXAMPLE VI

Investigation of PCR Bias using qPCR

[0187] The amplification bias associated with three DNA oligos was investigated using quantitative PCR (qPCR). The sequences of the three DNA templates were as follows:

Template A (SEQ ID NO:2)

ACAGGTTCAGAGTTCTACAGTCCGACGATCAGCTNNNNNNNNNNN
NNNNNTGGTGGAGCTGGCGGGAGTTGAACCCGCGTCCGAAATTCC
TACATCCTCGGTACTACATGGCCGTCGTATGCCGTCTTCTGCTTG

Template I (SEQ ID NO:30)

ACAGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNN
NCCGGGACCGTCCGAGCTTCGGATACCTAGACAAGCAGAAGACGG
CTGACCCTGTTTTCAGGGTCGCCGTCGTATGCCGTCTTCTGCTTG

Template J (SEQ ID NO:31)

ACAGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNNNNNNNN
NCGTAGTACCGTCCGGACCCTGTTTTCAGGGTCCGGACGGGCACAG
ATTAGCACCCTATCGACGAGCCGTCGTATGCCGTCTTCTGCTTG

[0188] The amplification efficiency of the three DNA templates was measured using qPCR. Primer 1 (SEQ ID NO:10) and Primer 2 (SEQ ID NO:11) were added to 1x Fast SYBR Green Master Mix (Applied Biosystems) to a final concentration of 0.5 μM. One of the three DNA template samples was then added to the PCR reaction mixture to one of four final concentrations (0.1 pM, 1 pM, 10 pM, or 100 pM). Amplification efficiencies were obtained for each reaction mixtures using a Fast Real-Time PCR 7500 System (Applied Biosystems) under the following thermocycling conditions:

1) 98 C for 20 s
2) 98 C for 3 s
3) 60 C for 30 s
4) Repeat steps 2-3 59 times

The resultant amplification efficiencies are shown in Figure 9.

EXAMPLE VII

Generation and Optimization of Barcodes

[0189] A set of 2,358 random 20-base optimized barcodes having a distance of nine was prepared using a computer such that even if a barcode accumulated nine mutations, it would not take the sequence of any other generated barcode sequences. Suitable software programs are commercially available that can be used to generate optimized barcode sequences *in silico* such as Python version 3.2.1 (world wide web python.org) and gcc compiler for C version 4.1.2 (world wide web gcc.gnu.org)

which were used for the generation and filtering of the barcode candidates described below.

[0190] A first barcode candidate containing 20 nucleotides was randomly generated *in silico*. Then, a second 20 nucleotide potential barcode candidate was randomly generated *in silico* and the number of sequence mismatches required to regenerate the first barcode candidate from the potential candidate (threshold value) was determined. If the threshold value is greater than 9 (i.e., the distance), the potential candidate is kept and added to the set of final barcode candidates; if not, the potential candidate is discarded and new potential candidates are generated randomly until the threshold value between the new potential candidate and the first barcode candidate is greater than 9; this new potential barcode candidate is then kept and added to the set of final barcode candidates. Subsequent sequences are then generated randomly and compared to all previously kept barcode candidates and discarded if allowing 9 or fewer sequence mismatches will allow the new potential barcode candidate to exactly take on the sequence of any previously kept barcode candidate. This process was repeated until we achieved a set of 2,358 final barcode candidates. It is to be understood that one can use this procedure on barcode candidates of any length or multiple lengths and with any threshold value criterion. Running this procedure multiple times should not result in the exact same set of barcode candidate sequences between runs.

[0191] One is also able to systematically generate barcode candidates by using the Hadamard Code as described in Bose RC, Shrikhande SS (1959) A note on a result in the theory of code construction. *Information and Control* 2:183-194 hereby incorporated by reference herein in its entirety, which allows generation of a set number of barcode candidates each a set, specific threshold value (number of allowed sequence mismatches required for a barcode candidate to exactly match the sequence of another barcode candidate) apart from all other barcode candidates. However, this technique is limited in that it requires that the length of the barcode be a power of 2.

[0192] Barcode candidates containing homopolymers longer than four bases or GC-content less than 40% or greater than 60% were discarded. Barcode candidates were also discarded if each exceeded a certain degree of complementarity or sequence identity (total matches and maximum consecutive matches) with (1) the Illumina paired-end sequencing primers described in Bentley, Nature 456:53-59 (2008) hereby incorporated by reference in its entirety, (2) the Illumina PCR primers PE 1.0 and 2.0, (3) the 3' end of the Illumina PCR primers PE 1.0 and 2.0, (4) the whole *E. coli* genome [K-12 MG1655 strain (U00096.2)], and (5) all other generated barcode candidates. Any barcode candidate for which an indel mutation would place it within five point mutations of another barcode candidate was also discarded. The final population consisted of 150 barcodes the sequences of which were identified, of which 145 were randomly chosen and used. It is to be understood that the present disclosure is not limited to specific sequences of barcodes within a set as the design of each set is likely to vary depending on the criteria used to identify barcodes and the methods to create them *in silico*.

[0193] Specifically, each of the 2,358 barcodes was analyzed for sequence characteristics that would contribute to either amplification or sequencing errors as follows.

[0194] ***Initial Filtering.*** All barcodes with less than 40% or greater than 60% GC content or containing homopolymers greater than length four were deleted. All barcode sequences were compared to the PE 1.0 and 2.0 Illumina PCR primer sequences and discarded if there were more than 10 total base matches or more than five consecutive base matches in any possible alignment with either primer sequence (sense and antisense). Each barcode was compared to the final four, five, and six bases closest to the 3' end of the PCR primer sequence (sense and antisense), respectively. The final six bases for both the PE 1.0 and PE 2.0 are identical. If any of these regions contained more than three consecutive base matches with a given barcode in any possible alignment, that barcode was discarded. All barcode sequences were compared to all other barcode sequences (including cases of offset sequences), and a barcode was discarded if there were more than 15 total base matches or 10 consecutive base matches to any other barcode sequence (sense and

antisense) in any possible alignment. The total number of hydrogen bonds in the longest consecutive matching region of each barcode to any other barcode sequence was calculated, and barcodes with greater than 26 total hydrogen bonds in that region were discarded. Each barcode was aligned with the entire *E. coli* genome (sense and antisense). If at any position in the genome a barcode contained more than 16 total matching bases, 12 consecutive matching bases, or 32 hydrogen bonds present in any given consecutively matching region, that barcode was discarded. Finally, all possible indels were generated for each barcode and the resulting sequence was compared to each original barcode sequence. If the resulting indel sequence could incur less than five point mutations and result in the exact sequence of any barcode, the barcode sequence that generated the given indel was deleted.

[0195] ***Score Filtering.*** In addition to the thresholds described above for characterizing an optimized barcode set, a more in-depth analysis of barcode-barcode and barcode-*E. coli* genome hybridization was performed particularly regarding barcode hybridization melting temperatures with respect to PCR amplification. Although for a given barcode sequence, when comparing complementarity to a large set of reference sequences through sequence alignment there will be an alignment condition which results in a region in the barcode sequence where the absolute maximum number of consecutive base matches is achieved (as described above); there are other possible conditions when a barcode contains a region where the number of consecutive bases in the region of maximum consecutive base matches does not reach the absolute maximum value as described above. This value for any given region is referred to as the score. When comparing a barcode to the sense and antisense sequences of all other barcodes, all alignment conditions were determined in the cases where the score of the region that contained the highest number of consecutive base matches (first score) were 10, nine, eight, and seven bases, respectively. For each of these four first scores, the condition where the maximum score of the region where the second-highest number of consecutive base matches (second score) occurred was determined. For example, a condition where the first score is 10 and the second score is three is denoted as 10-3 and this condition is defined as the duplex. If the sum of the first score and the second score compared to

all other barcode sequences for any barcode was greater than 12, that barcode was discarded. If the distance between these two regions (maximum and second-most consecutive matches) was one base, and the sum of the first score and the second score was greater than 11, the barcode was discarded. The maximum value of the number of consecutive base matches for a region under all alignment conditions that contained the third-highest number consecutive base matches (third score) was also determined for all barcodes. Given the maximum third score, the respective maximum first score was determined; the maximum second score given both of these conditions was also determined. This condition is defined as the triplex and denoted, for example, as 7-4-3. Barcodes with the following triplexes were manually deleted: 7-3-3, 6-5-4, 6-5-3, 6-4-4, 5-5-4, and 5-4-4. Barcodes with a triplex of 6-4-3 were deleted where both the distances between adjacent regions corresponding to the scores was one base.

[0196] The same analysis was done for all barcodes aligned against the entire *E. coli* genome (sense and antisense). Barcodes with a first score and second score sum of greater than 15 were discarded. Barcodes where the first score region and the second score region were separated by one nucleotide and had a first score and second score sum of more than 14 were discarded as well. Barcodes with the following triplexes were deleted: 8-4-4, 7-5-4, 6-6-4, 6-5-5, and 5-5-5. After filtering, a total of 150 barcodes remained.

EXAMPLE VIII

In-Depth Design and Preparation of Adapter

[0197] ***Adapter Design.*** To avoid sequencing errors resulting from cluster overlap (i.e. low sequence complexity) and to reduce potential ligation bias, an additional two to five base extension—CT, ACT, GACT, or TGACT—was added to the 3'-end of each barcode. These sequences mimic the T-overhang in the conventional Illumina paired-end adapter and conserve the sequence of the last two bases. For each of the 150 final barcodes, these four different adapter extensions were attached to the 3' end of the barcode. The same values as used in the initial filtering step (see above) were obtained for each of the four adapter candidates for all 150 barcodes. The following

four parameters of analysis were determined: PCR primer matching (PC), 3' end of PCR primer matching (TP), barcode-barcode matching (BB), and barcode-*E.coli* genome matching (EC) and calculated the complementarity score of each category for all barcode-adapter candidates as follows: PC = {Sum of [maximum total base matches to the PE 1.0 and PE 2.0 PCR primers (sense and antisense for a total of four terms)]} + 2·{Sum of [maximum consecutive base matches to the PE I.0 and PE 2.0 PCR primers (sense and antisense for a total of four terms)]^2}; TP = {Sum of [maximum total base matches to the final four bases of the PCR primer sequence (sense and antisense for a total of two terms)]^2} + 1.5·{Sum of [maximum total base matches to the final five bases of the PCR primer sequence (sense and antisense for a total of two terms)]^2} + 2·{Sum of [maximum total base matches to the final six bases of the PCR primer sequence (sense and antisense for a total of two terms)]^2}; BB = {Sum of [maximum total base matches to all other barcode candidates (sense and antisense for a total of two terms)]^2} + 2·{Sum of [maximum consecutive base matches to all other barcode candidates (sense and antisense for a total of two terms)]^2}; EC = Maximum total base matches to entire *E. coli* genome (sense only) + 2· [maximum consecutive base matches to the entire *E. coli* genome (sense only)]^2.

[0198] The total complementarity score (TC) for each barcode candidate was calculated as follows: TC = 3·PC + 15·TP + BB + EC. The TC value gives a metric to determine the expected efficacy of each barcode candidate during PCR amplification. A low TC represents a lower chance of amplification errors caused by unwanted hybridization between barcodes and adapters, primers, or the sample. For each barcode, the barcode-adapter candidate was selected that had either the lowest or second lowest TC among the four. This resulted in 150 final barcode-adapter sequences, of which 145 were randomly chosen and used. 37 CT extensions and 36 of each of the other three extensions were used. Adapters were then designed in the same Y-shaped construct as the conventional Illumina paired-end adapter with a 22 to 25 base-pair extension that contained the barcode and a T-overhang (Fig. 10*B*). Both strands (A and B) of the adapter were ordered from Integrated DNA Technologies (IDT).

[0199] ***Adapter Generation.*** The 5'-end of strand B was phosphorylated in T4 DNA Ligase Reaction Buffer (New England Biolabs, (NEB)) containing 40 μM strand B and 20 U T4 polynucleotide kinase (NEB) at 37 °C for 60 min in 20 μL, followed by a 25 minute incubation at 70 °C, and a 5 minute incubation at 90 °C for enzyme inactivation. The phosphorylated strand B was annealed to each respective strand A in NEB Buffer 2 (NEB). Each solution contained 20 μM strand A and 20 μM strand B in a total volume of 20 μL. The solutions were first raised to 90 °C and cooled to 25 °C at a rate of 5 °C /minute (Annealing Temperature Condition). Finally, equal volumes of all 145 annealed adapters were mixed.

EXAMPLE IX

Design and Preparation of Spike-in and Normalization DNA

[0200] 15,000 random 30 base-pair sequences were generated such that even if a sequence accumulated 15 mutations, it would still be identifiable and distinguishable from all other generated sequences. Spike-in and normalization candidates with a maximum homopolymer length of greater than 3, or a GC-content less than 11 or greater than 19 were discarded. Spike-in and normalization candidates were also discarded if they exceeded a certain degree of complementarity or sequence identity (total matches and maximum consecutive matches) with (1) the Illumina paired-end sequencing primers, (2) the 3'-end of the sequencing primers, (3) the whole *E. coli* genome [K-12 MG1655 strain (U00096.2)], and (4) all other generated spike-in candidates in the same fashion as barcode design. The final population consisted of 40 spike-in and normalization DNA candidates, of which three were chosen at random seven times (without replacement) and concatenated, with one deletion at the 60^{th} base of strand A (corresponding to the 31^{st} base of strand B) and an addition of a single A to the end of the sequence to form seven 90-base spike-in DNA sequences and one normalization DNA sequence. Both strands of 5'-end phosphorylated DNA oligos were ordered from IDT and were annealed in 0.3x NEB Buffer 2 (NEB) with 50 μM of each strand using the Annealing Temperature Gradient. All seven spike-in sequences were ligated to the barcoded adapter mixture in NEBNext Quick Ligation Reaction Buffer (NEB) with 6.7 μM annealed spike-in, 6.7 μM barcoded adapter, and 6 μM Quick Ligase (NEB) by incubating at 25 °C for 30 min. The product was run

on a 5% polyacrylamide gel (Bio-Rad), and the targeted band (at ~270 bp) was removed from the gel. The gel slice was cut into small pieces and the embedded DNA was extracted into diffusion buffer (10 mM Tris-Cl pH 8.0, 50 mM NaCl, 0.1 mM EDTA) by overnight incubation at room temperature. Then, the extracted spike-ins were purified on a column (Zymo Research). Sequence analysis (GeneWiz) confirmed that the band contained the expected ligation product. The concentration of each spike-in was estimated by qPCR (Fast SYBR Master Mix, Applied Biosystems) using sequence-specific qPCR primers against a known-concentration Y-shaped Reference DNA (below). The concentrations of spike-ins for the second deep sequencing run were measured in parallel by digital PCR (Fluidigm) at Molecular Genetics Core Facility of Children's Hospital Boston Intellectual and Developmental Disabilities Research Center. Each spike-in was measured a total of ten times on two separate chips (48.770).

EXAMPLE IX

Design and Preparation of Y-shaped Reference DNA

[0201] From the original list of 150 barcode candidates, we chose two barcodes that were not present in the final list of 145 used. Then we concatenated the two barcodes with the Y-shaped adapter sequences and a 90-base pair targeted sequence mimic such that the targeted sequence mimic was between the barcodes, which were between the adapters. The 90-base pair targeted sequence mimic was designed the same way as the spike-in and normalization DNAs. Both strands of DNA oligos were ordered from IDT and their concentrations were measured by absorbance at 260 nm using the extinction coefficient provided by IDT. The DNA oligos were annealed in water with 5 μM of each strand using the Annealing Temperature Gradient.

EXAMPLE X

E.coli RNA Preparation and cDNA Generation

[0202] *E. coli* [K-12 MG1655 strain (U00096.2)] was grown overnight at 30 °C in LB medium. The resulting culture was diluted 500-fold in fresh LB medium and grown at 30 °C for 3.5 hours such that O.D. at 600 nm became 0.30-0.35. 1 mL of cells was quickly killed by addition of 0.1 mL stop solution (90% (v/v) Ethanol and 10% (v/v)

Phenol). The cells were collected by centrifugation (9,100 x g, 1.5 min, room temperature), suspended in 1 mL cooled PBS (Lonza), and centrifuged again (16,000 xg, 1.5 min, room temperature). The supernatant was removed and the cells were suspended in 0.1 mL of 1 mg/mL lysozyme in TE Buffer (pH 8.0) (Ambion). 0.1 mL of lysis buffer (Genosys) was added and the mixture was vortexed for 5 s. After adding 0.2 mL of Phenol Chloroform pH 4.5 (Sigma) and vortexing three times for 5 s, the mixture was centrifuged (16,000 x g, 3 min, room temperature). The top layer of solution was taken and 0.15 mL of 100 % 2-Propanol (Sigma) was added; the mixture was left on ice for 30 min. The solution was centrifuged (16,000 x g, 30 min, 4°C) to precipitate the RNA. The RNA pellet was washed twice by centrifugation (16,000 x g, 5 min, 4°C) with 0.75 mL of cold 70% (v/v) ethanol. After the second centrifugation, the supernatant was removed and the pellet was dried for 15 min at room temperature. Then, 88 μL of water was added and the mixture was incubated for 15 min at room temperature, followed by resuspension. The resulting solution was mixed with 0.04U/μL DNase I (NEB) in DNase I Reaction Buffer (NEB) for a total volume of 100 μL, and the mixture was incubated at 37 °C for 30 min followed by addition of EDTA (Sigma) to a final concentration of 5 mM. The mixture was incubated at 75 °C for 10 min to inactivate DNase I, followed by column purification. Ribosomal RNA was removed using Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria) (Epicentre, Illumina). From this point, the conventional Illumina protocol for mRNA Sequencing Sample Preparation was followed with a few modifications. The purified RNA was fragmented in 0.5x fragmentation buffer (Ambion) with ~500 ng RNA in a 100 μL reaction solution. The solution was incubated on ice for 1 min after the fragmentation buffer was added followed by a 6 min incubation at 70 °C. The tube was placed on ice and incubated for 1 min followed by addition of 4 μL stop solution (Ambion). The fragmented RNA was purified with a column and eluted in 11.1 μL in water. 1 μL of 50 μM Random Hexamer Primer (Applied Biosystems) was added to this solution and incubated at 65°C for 5 minutes and then placed on ice. 4 μL 5x First Strand Buffer (Invitrogen), 2 μL 100 mM DTT (Invitrogen), 0.4 μL 25 mM dNTP Mix (Applied Biosystems), and 0.5 μL RNase inhibitor (Applied Biosystems) was added to the mixture. This was incubated at 25°C for 2 minutes, followed by the addition of 1 μL

Superscript II (Invitrogen). It was then incubated at 25 °C for 10 minutes, 42°C for 50 minutes, and 70°C for 15 minutes to synthesize the first strand of the cDNA and inactivate the enzyme, which was placed on ice and then purified on a column. The eluate from this column was used to generate the second strand of cDNA in NEBNext Second Strand Synthesis Reaction Buffer (NEB) with 0.3 U/μL DNA polymerase I (*E. coli*) (NEB), 1.25 U/μL *E.coli* DNA Ligase, and 0.25 U/μL RNase H in an 800 μL total volume solution at 16 °C for 2.5 hours followed by the column purification. The eluted double stranded cDNA was end-repaired in T4 DNA Ligase Buffer (NEB) with 0.4 mM Deoxynucleotide Solution Mix (NEB), 0.5 U/μL T4 DNA polymerase (NEB), 0.5 U/μL T4 Polynucleotide Kinase (NEB) in a 200 μL reaction solution by incubating at 20 °C for 30 min followed by column purification. The eluted end-repaired cDNA was dA-tailed in NEB 2 buffer (NEB) with NEBNext dA-tailing Reaction Buffer with 1mM dATP (NEB) and 0.3 U/μL Klenow Fragment (3' →5' exo-) in a 50 μL solution by incubating at 37 °C for 30 min followed by column purification.

EXAMPLE XI

Sample-Adapter Ligation, Sequencing Sample Preparation, and Sequencing

[0203] The cDNA library was ligated to the barcoded adapter mixture and conventional Illumina paired-end adapter (without phosphorothioate bond) (IDT) respectively in the NEBNext Quick Ligation Reaction Buffer (NEB) with 5.4 μL of the cDNA produced above, 1.9 μM barcoded adapter or conventional Illumina Paired-end adapter, and 3.6 μM Quick Ligase (NEB) for a total volume of 10 μL by incubating at 25 °C for 15 min. The two solutions were separately run on a 5% polyacrylamide gel and the portion between 250-300 bp was cut. The gel slice was cut into small pieces and the embedded DNA was extracted into diffusion buffer (10 mM Tris-Cl pH 8.0, 50 mM NaCl, 0.1 mM EDTA) by overnight incubation at room temperature. Then, the extracted DNAs were column purified. The concentrations of purified products were measured by qPCR (Fast Fast SYBR Master Mix, Applied Biosystems) against a known-concentration Y-shaped reference sequence using designed qPCR primers purchased from IDT. The sample ligated to the barcoded adapter and the conventional Illumina Paired-end adapter were amplified by PCR (1

cycle of 98 °C for 1 min, 18 cycles of 98 °C for 1 s, 65 °C for 45 s, 72 °C for 40 s, and 1 cycle of 72 °C for 5 min) in HF buffer (NEB) with 0.63 mM dNTP, 0.5 mM of each amplification primer modified from Illumina PCR primer PE 1.0 and 2.0 (IDT), 25 fM DNA sample, and Phusion DNA polymerase (NEB) in 20 μL, with spike-in DNAs (0.71 aM Spike-in 1, 1.0 aM Spike-in 2, 4.4 aM Spike-in 3, 18 aM Spike-in 4, 150 aM Spike-in 5, 480 aM Spike-in 6 for the first sequencing run, and 1.3 aM Spike-in 1, 6.9 aM Spike-in 3, 36 aM Spike-in 4, 150 aM Spike-in 5, 770 aM Spike-in 7 for the second sequencing run). Then, 10 pM Normalization DNA was added to both PCR products, and the DNA was purified twice on a column. The concentration of the purified product was measured by qPCR (Fast Fast SYBR Master Mix, Applied Biosystems) using the conventional Illumina qPCR primer (IDT) against a PCR product amplified from Y-shaped reference DNA using modified Illumina PCR primers whose concentration was measured by NanoDrop (LMS). The final concentration of each spike-in and normalization DNA in the purified products was measured by qPCR using sequence-specific qPCR primers described above and compared by normalization. The length distribution of the purified PCR product was measured by Bioanalyzer (Agilent). Samples with barcoded adapters were sequenced on an Illumina HiSeq 2000 with 2x100 (for the first sequencing run) and 2x50 (for the second) base paired-end reads in one lane.

EXAMPLE XII

Quantification Accuracy Using Digital PCR and Digital RNA-Seq of Spike-ins

[0204] ***Spike-in Analysis.*** From the raw sequencing data, reads were isolated which contained barcode sequences that corresponded to the original 145 barcodes in both forward and reverse reads for each sequencing cluster that had at most one mismatch. The first 28 bases (26 bases for the second sequencing run) of the targeted sequence of both the forward and reverse reads of each cluster were aligned to each Spike-in sequence, which is known. Sequences with more than two mismatches were discarded. The number of unique tags present in each spike-in were counted to determine the number of copies of each spike-in.

[0205] To calibrate the digital RNA-Seq system, the concentrations of five synthetic DNA spike-in sequences were measured using the Fluidigm digital PCR platform and used as internal standards. The spike-in samples were barcoded, added to the barcoded *E. coli* cDNA library, and quantified using the sequencing-based digital counting strategy described above. Fig. 11*A* shows that the number of digital counts (i.e. unique barcodes) observed in deep sequencing is well-correlated with the digital PCR calibration of the spike-in sequences.

[0206] To evaluate the difference between using random barcode sequences and optimized barcode sequences, two experiments were conducted. In one experiment, the spike-in molecules were labeled with random barcode sequences, and in the second experiment, the optimized, pre-determined barcode set was used. The histograms of the number of reads for all barcodes observed from the most abundant spike-in sequence were constructed (Fig. 11*B*). When using random barcodes (light histogram in Fig. 11*B*, the left-most bin exhibits a large peak because a substantial fraction of barcodes are infrequently read due to sequencing errors. This causes barcodes to interconvert, generating quantification artifacts. In stark contrast, the left-most bin when using optimized barcodes (dark histogram in Fig. 11*B*) has no such peak because the optimized barcode sequences avoid misidentification due to sequencing errors. The effect of sequencing error on both random and optimized barcode counting is clearly shown by simulation (Fig. 12).

[0207] The dark histogram in Fig. 11*B* is the distribution of the number of reads for the 5,311 uniquely barcoded molecules from a particular spike-in. Assuming each barcoded spike-in molecule is identical, the dark histogram in Fig. 11*B* is essentially the probability distribution of the number of reads for a single molecule, which spans three orders-of-magnitude. This broad distribution arises primarily from intrinsic PCR amplification noise in sample preparation. Given this broad single molecule distribution, for low copy molecules in the original sample, counting the total number of reads (conventional RNA-Seq) would lead to inaccuracies. On the other hand, this problem can be circumvented if one counts the number of different barcodes (integrated area of the histogram) using the digital RNA-Seq approach, yielding

accurate quantification with single copy resolution. The two counting schemes give same results only when the copy number in the original sample is high, assuming there is no sequence-dependent bias. Random sampling of the barcode sequences by each target sequence is essential for accurate digital counting. Fig. 11*C* shows that the distribution of observed molecule counts is in excellent agreement with Poisson statistics. Therefore the five spike-in sequences sample the 21,025 barcode pairs without bias.

EXAMPLE XIII

Efficacy of Downsampling of Spike-in Reads

[0208] The spike-in reads in the replicate experiment were randomly downsampled by a factor of 10. For each read of each of the five spike-ins, there was a 10% chance that it was kept and counted, whereas the other 90% was discarded. Fig. 13A shows that there is little dropout between these two conditions (the data show the average dropout rate for the five spike-ins to be 1.6%) and that for the spike-in with the highest number of molecules, the change in the single molecule coverage histogram is minimal.

EXAMPLE XIV

Digital Quantification of the E. coli Transcriptome

[0209] ***E. coli* Transcriptome Analysis.** Reads were isolated which contained barcode sequences that corresponded to our original list of 145 barcodes in both forward and reverse reads for each sequencing cluster that had at most one mismatch. The first 28 bases (26 bases for the second sequencing run) of the targeted sequence of both the forward and reverse reads of each cluster were aligned to the *E. coli* genome and the sequences that uniquely align fewer than three mismatches and where the two reads did not map to the same sense or antisense strand of the genome were kept. The remaining sequences were mapped to transcription units as described in Keseler, Nucleic Acids Res 39(Database issue):D583-D590 (2011) and sorted by starting and ending position as well as forward and reverse barcodes (unique tag). Mapped sequence fragments with a length of at least 1,000 bases were discarded. All

sequences within the same transcription unit that had the same unique tag were analyzed further. It was determined that more than one sequence with the same unique tag were identical if the distance between their center positions was less than four base-pairs and if the difference in length was less than 9 base-pairs (Fig. 14 and Fig. 15). Thus, the read counts for sequences deemed identical were summed and the sequence with more read counts was deemed as the actual correct sequence. Then for each unique sequence, the number of unique barcode tags that appeared to determine the copy number of each sequence were counted.

[0210] Comparison of Noise in Conventional vs. Digital Counting for the *E. coli* Transcriptome. The total number of reads and the total number of digital counts for each base in each of the mapped sequences were summed. For each transcription unit, bins were created that were 99 base-pairs long and the total number of reads and the total number of digital counts present in each bin were summed. Bins that yielded an average number of digital counts per base of greater than or equal to 1 were selected and from these bins, the average and sample standard deviation of the summed reads and summed digital counts were counted, respectively. The noise was defined as the sample standard deviation divided by the mean and this value was calculated for both reads and digital counts; the ratio of the noise for reads to digital counts was computed for each transcription unit (Fig. 16*D*).

[0211] We obtained 26-32 million reads from the barcoded cDNA libraries that uniquely mapped to the *E. coli* genome in two replicate experiments. Fig. 16*A* shows the number of conventional and digital counts (unique barcodes) as a function of nucleotide position for the fumA transcription unit (TU). Not surprisingly, the read density is considerably less uniform across the TU than the number of digital counts, presumably due to intrinsic noise and bias in fragment amplification.

[0212] It is advantageous for transcripts across the *E. coli* transcriptome to sample all barcodes evenly. Fig. 16*B* shows this distribution, which is close to Poisson but is somewhat overdispersed. Such biased sampling reduces the effective number of barcode sequences N_{eff} available. However, in the *E. coli* transcriptome sample, the

copy number of the most abundant cDNA ranges from 10-40 copies for both counting methods. Based on Poisson statistics, even for the most abundant cDNA fragments in our sample, the required N_{eff} is ~100-400 for 95% unique labeling of all molecules. Because there are 21,025 barcode pairs available, on average the degree of randomness observed in Fig. 16*B* is sufficient.

[0213] Conventional methods count the number of amplicons, a quantity that is subject to bias and intrinsic amplification noise, rather than the number of molecules in the original sample. Conversely, in the digital counting scheme described herein, unique barcode sequences distinguish each molecule in the sample, and so the effects of intrinsic noise are minimized, especially when an optimized barcode set is used. Fig. 16*C* shows how drastically different digital counting can be from conventional counting at low copy numbers, implying that digital counting of unique barcodes is advantageous, particularly for quantifying low copy fragments. The correlation is stronger for high copy fragments and the same phenomenon is also observed for whole TUs and genes (Fig. 17).

[0214] To demonstrate the superior accuracy of digital counting, the uniformity of abundance measurements within individual transcripts was examined. Because individual TUs were, by-and-large, intact RNA molecules following RNA synthesis, the cDNA fragments that map to one region of a given TU should have the same abundance as fragments that map to a different region of the same TU. The ratio between the variation in conventional counting ν_C and variation in digital counting ν_D for TUs in different abundance ranges were histogrammed (Fig. 16*D*). A variation ratio of $\nu_C/\nu_D = 1$ indicates that both conventional and digital counting give similarly uniform abundances along the length of a TU. For a TU where ν_C/ν_D exceeds one, conventional counting measures abundance less consistently along the TU than digital counting. The mean values of ν_C/ν_D in the two replicates are 1.4 (s=1.5, where s is sample standard deviation) and 1.2 (s=0.5) for the complete set of analyzed TUs, indicating that conventional counting is less consistent than digital counting across an average TU. Furthermore, the mean value of ν_C/ν_D increases with decreasing copy number and its distribution becomes broader (Fig. 16*D*). For TUs in

the lowest abundance regime, the mean values of ν_C/ν_D are 1.9 (s=2.4) and 1.3 (s=0.9) for the two replicates. On average, digital counting outperforms conventional counting in terms of accuracy, and its performance advantage is most pronounced for low abundance TUs.

[0215] While Fig. 16 demonstrates that digital counting is less noisy and more accurate than conventional counting, Fig. 18 shows that digital counting is also more reproducible. This is demonstrated on the level of a single TU in Fig. 18*A*, which shows the ratio of counts between the two replicates for both conventional and digital counting along the fumA transcript. This ratio is consistently close to one for digital counting, but fluctuates over three orders-of-magnitude for conventional counting. We analyzed the global reproducibility of the whole transcriptome for quantification of TUs and genes for both conventional and digital counting in Fig. 18*B* and Fig. 18*C*, respectively. In both cases, the correlation between replicates is noticeably better for digital counting than conventional counting, particularly for low copy transcripts.

[0216] Fig. 19 is a simulation demonstrating the advantageous performance of digital counting methods described herein over conventional counting in differential expression analysis. RNA expression quantification was simulated using experimentally measured copy numbers, barcode sampling, and amplification noise distributions for two different libraries for each of three different systems (*E. coli* transcriptome fragments in Figure 19A, *E. coli* transcription units in Figure 19B, and human stomach microRNA in Figure 19C. The ratio of simulated to actual fold-change for each gene as a function of the lower of two copy numbers for the two compared libraries is plotted. Ideally, the value of this ratio is one for all genes. Because digital counting is almost completely immune to amplification noise, it gives consistently superior performance to conventional counting for differential expression, even at low copy numbers. The discrepancy between conventional and digital counting is smaller for the *E. coli* transcription unit library in (*B*) than for the fragment library in (*A*) because amplification noise can be averaged over many fragments in the case of long transcription units.

EXAMPLE XV

Digital Quantification of Copy Numbers of Different Sequences at Different Concentrations

[0217] According to an aspect of the present disclosure, methods are provided for counting copy numbers of different sequences that are present in a sample at different concentrations using unique barcodes. Unlike conventional methods for counting molecules with very different copy numbers, the methods described herein using unique barcodes allows the entire sample to be processed in a single tube.

[0218] As a precursor, consider the conventional method which uses a dilution scheme requiring many tubes. In this implementation of molecular counting, one is able to combine sample dilutions by attaching the same barcode to all molecules in the same tube for all tubes to achieve digital counting. Consider the Path A in Figure 20 in which a sample containing two species (DNA 1 and DNA 2) which have copy numbers of 1111 and 11, respectively. After performing order-of-magnitude serial dilutions several times, the sample is now split into five different tubes, where each sequential tube contains 10-fold fewer copies of each DNA than the previous tube (Path A). The tubes are numbered from the left as tube 0, tube 1, etc. During the serial dilution process, at some point a given tube (tube 2) will no longer contain any copies of DNA2 but will still contain 10 copies of DNA1 due to a much higher number of DNA1 molecules in the original sample. The final tube (tube 4) has been diluted such that both DNA1 and DNA2 are no longer present. All molecules in a given tube are labeled with the same barcode sequence and the procedure repeated for all tubes such that each different tube has a unique barcode sequence which maps to the specified tube. For example, all DNA molecules from tube 0 will be labeled by a specific barcode, whereas all DNA molecules from tube 1 will be labeled by a different barcode as illustrated.

[0219] After this process, all molecules from all tubes are combined together. Despite mixing together the DNA from all tubes, the barcoding scheme still allows tracking of which tube a given molecule is originally from. The combined sample will then be subjected to PCR amplification and then massively parallel sequencing; conventional PCR can easily be implemented by designing barcodes which contain the amplification primer sequences. After sequencing, four types of barcodes are detected for DNA1: barcode 1

(tube 0), barcode 2 (tube 1), barcode 3 (tube 2), and barcode 4 (tube 3). However, the barcode 5 representing dilution by 10^4 (tube 4) does not appear. This suggests that the copy number of DNA1 originally present is at least (and is the same order-of-magnitude as) $10^0 + 10^1 + 10^2 + 10^3 = 1{,}111$, which is the same as the known original copy number of DNA1. Similarly, only two types of barcodes are detected for DNA2: barcode 1 (tube 0) and barcode 2 (tube 1), which suggests that the copy number of DNA2 originally present is at least (and is the same order-of-magnitude as) $10^0 + 10^1 = 11$.

[0220] The same process may be applied to the same sample using only one tube and without any serial dilutions by utilizing stochastic labeling of DNA molecules by barcodes. Consider now the same original tube in Figure 20 which contains 1111 DNA1 molecules and 11 DNA2 molecules. Instead of making serial dilutions, we will instead add a barcode set to the original tube (Path B), which contains five barcodes in the ratio 10,000:1,000:100:10:1. By adding the entire barcode set to the original sample tube, one would expect that labeling of all copies of both DNA molecules by said barcodes would also follow the same ratio. After barcode labeling and PCR amplification, the original tube is subjected to massively parallel sequencing, and the results are exactly the same as in Path A; there were at least 1111 DNA1 molecules and at least 11 DNA2 molecules. In effect, this labeling procedure is identical to the dilution protocol presented above. According to method B, digital counting is performed in a single tube by using a small number of barcodes (only one per order-of-magnitude) on a large number of unique DNA molecules with highly varying copy numbers. Accordingly a method is provided of determining the copy number of different nucleic acids in a sample by adding to the sample a set number of barcodes in a ratio to one another, allowing the barcodes to attach to the nucleic acids according to the ratio and, sequencing the nucleic acids to determine the copy number. Accordingly a method is provided of determining the copy number of different nucleic acids in a sample by attaching barcodes to the nucleic acids where the barcodes have an order of magnitude ratio to one another and, sequencing the nucleic acids to determine the copy number. The methods of attaching barcodes to nucleic acids in a sample, amplifying and sequencing as describe herein can be used with the methods of this Example XV.

[0221] Both methods A and B have probabilistic uncertainty and these uncertainties are essentially the same. One can decrease this uncertainty dramatically and as a result accurately count all molecules on average by performing multiple replicate experiments. A consideration of this technique is the copy number of the highest copy barcode, which must be higher than the copy number of the most abundant DNA species in the original tube. Otherwise, there may not be enough resolution to determine the exact copy number of the higher copy templates. Although relatively low resolution is required for the example in Figure 20, as the ratio of barcode copy numbers is separated by an order of magnitude each, it is very plausible to have many more barcodes each separated by a factor of two or less. Although this would increase the number of barcodes required, precision is also increased. To cover a 10^6-fold range, one only needs 20 barcodes, assuming the case of two-fold separation. With this method, one can easily determine the copy number of multiple nucleic acid templates within a single tube without excessive or complicated sample preparation protocols.

What is claimed is:

1. A method of determining copy number of a nucleic acid molecule in a sample including a plurality of nucleic acid molecules comprising

attaching a unique barcode sequence to substantially each of the plurality of nucleic acid molecules in the sample to produce a plurality of barcoded nucleic acid molecules,

amplifying the plurality of barcoded nucleic acid molecules in the sample to produce amplicons of the plurality of barcoded nucleic acid molecules,

sequencing each amplicon to identify an associated nucleic acid sequence and an associated barcode sequence,

selecting a first target nucleic acid sequence and determining the number of unique associated barcode sequences for the first target nucleic acid sequence, wherein the number of unique associated barcode sequences is the copy number of the first target nucleic acid sequence.

2. The method of claim 1 wherein the nucleic acid molecules are DNA or RNA.

3. The method of claim 1 wherein the plurality of barcoded nucleic acid molecules is a plurality of barcoded RNA molecules and further including the steps of reverse transcribing the plurality of barcoded RNA molecules to produce barcoded cDNA molecules and amplifying the barcoded cDNA molecules to produce amplicons of the barcoded cDNA molecules.

4. The method of claim 3 comprising repeatedly reverse transcribing the plurality of barcoded RNA molecules to produce linear pre-amplified barcoded cDNA molecules and amplifying the linear pre-amplified barcoded cDNA molecules to produce amplicons of the linear pre-amplified barcoded cDNA molecules.

5. The method of claim 4 wherein the step of repeatedly reverse transcribing the plurality of barcoded RNA molecules includes using reverse transcriptase and a nicking enzyme.

6. The method of claim 1 wherein the plurality of barcoded nucleic acid molecules is a plurality of barcoded DNA molecules and further including the steps of repeated replication of the plurality of barcoded DNA molecules to produce a plurality of pre-amplified barcoded DNA molecules and amplifying the plurality of pre-amplified barcoded DNA molecules to produce amplicons of the plurality of pre-amplified barcoded DNA molecules.

7. The method of claim 6 wherein the step of repeated replication of the plurality of barcoded DNA molecules includes using DNA polymerase and a nicking enzyme.

8. The method of claim 1 wherein the sample is obtained from one or more cells of a first cell type and wherein amplification includes use of a primer generated from genomic DNA of the first cell type.

9. A method of counting nucleic acid molecules in a sample including a plurality of nucleic acid molecules comprising

attaching a unique barcode sequence to substantially each of the plurality of nucleic acid molecules in the sample to produce a plurality of barcoded nucleic acid molecules,

amplifying the plurality of barcoded nucleic acid molecules in the sample to produce amplicons of the plurality of barcoded nucleic acid molecules,

sequencing each amplicon to identify an associated barcode sequence, and

counting the number of unique associated barcode sequences as a measure of the number of nucleic acid molecules in the sample.

10. The method of claim 9 wherein the nucleic acid molecules are DNA or RNA.

11. The method of claim 9 wherein the plurality of barcoded nucleic acid molecules is a plurality of barcoded RNA molecules and further including the steps of reverse transcribing the plurality of barcoded RNA molecules to produce barcoded cDNA molecules and amplifying the plurality of barcoded cDNA molecules to produce amplicons of the plurality of barcoded cDNA molecules.

12. The method of claim 11 comprising repeatedly reverse transcribing the plurality of barcoded RNA molecules to produce linear pre-amplified barcoded cDNA molecules and amplifying the linear pre-amplified barcoded cDNA molecules to produce amplicons of the linear pre-amplified barcoded cDNA molecules.

13. The method of claim 12 wherein the step of repeatedly reverse transcribing the plurality of barcoded RNA molecules includes using reverse transcriptase and a nicking enzyme.

14. The method of claim 9 wherein the plurality of barcoded nucleic acid molecules is a plurality of barcoded DNA molecules and further including the steps of repeated replication of the plurality of barcoded DNA molecules to produce a plurality of pre-amplified barcoded DNA molecules and amplifying the plurality of pre-amplified barcoded DNA molecules to produce amplicons of the plurality of pre-amplified barcoded DNA molecules.

15. The method of claim 14 wherein the step of repeated replication of the plurality of barcoded DNA molecules includes using DNA polymerase and a nicking enzyme.

16. The method of claim 9 wherein the sample is obtained from one or more cells of a first cell type and wherein amplification includes use of a primer generated from genomic DNA of the first cell type.

17. A method of determining copy numbers of nucleic acid molecules in a sample comprising

attaching a unique barcode sequence to substantially each of the nucleic acid molecules in the sample to produce a plurality of barcoded nucleic acid molecules,

amplifying the plurality of barcoded nucleic acid molecules in the sample to produce amplicons of the plurality of barcoded nucleic acid molecules,

massively parallel sequencing the amplicons of the plurality of barcoded nucleic acid molecules to identify for each amplicon an associated nucleic acid sequence and an associated barcode sequence, and

determining the number of unique associated barcode sequences for each nucleic acid sequence in the sample.

18. The method of claim 17 wherein the nucleic acid molecules are DNA or RNA.

19. The method of claim 17 wherein the plurality of barcoded nucleic acid molecules is a plurality of barcoded RNA molecules and further including the steps of reverse transcribing the plurality of barcoded RNA molecules to produce barcoded cDNA molecules and amplifying the plurality of barcoded cDNA molecules to produce amplicons of the plurality of barcoded cDNA molecules.

20. The method of claim 19 comprising repeatedly reverse transcribing the plurality of barcoded RNA molecules to produce linear pre-amplified barcoded cDNA molecules and amplifying the linear pre-amplified barcoded cDNA molecules to produce amplicons of the linear pre-amplified barcoded cDNA molecules.

21. The method of claim 20 wherein the step of repeatedly reverse transcribing the plurality of barcoded RNA molecules includes using reverse transcriptase and a nicking enzyme.

22. The method of claim 17 wherein the plurality of barcoded nucleic acid molecules is a plurality of barcoded DNA molecules and further including the steps of repeated replication of the plurality of barcoded DNA molecules to produce a plurality of pre-amplified barcoded DNA molecules and amplifying the plurality of pre-amplified barcoded DNA molecules to produce amplicons of the plurality of pre-amplified barcoded DNA molecules.

23. The method of claim 22 wherein the step of repeated replication of the plurality of barcoded DNA molecules includes using DNA polymerase and a nicking enzyme.

24. The method of claim 17 wherein the sample is obtained from one or more cells of a first cell type and wherein amplification includes use of a primer generated from genomic DNA of the first cell type.

SUBSTITUTE SHEET (RULE 26)

FIG. 1

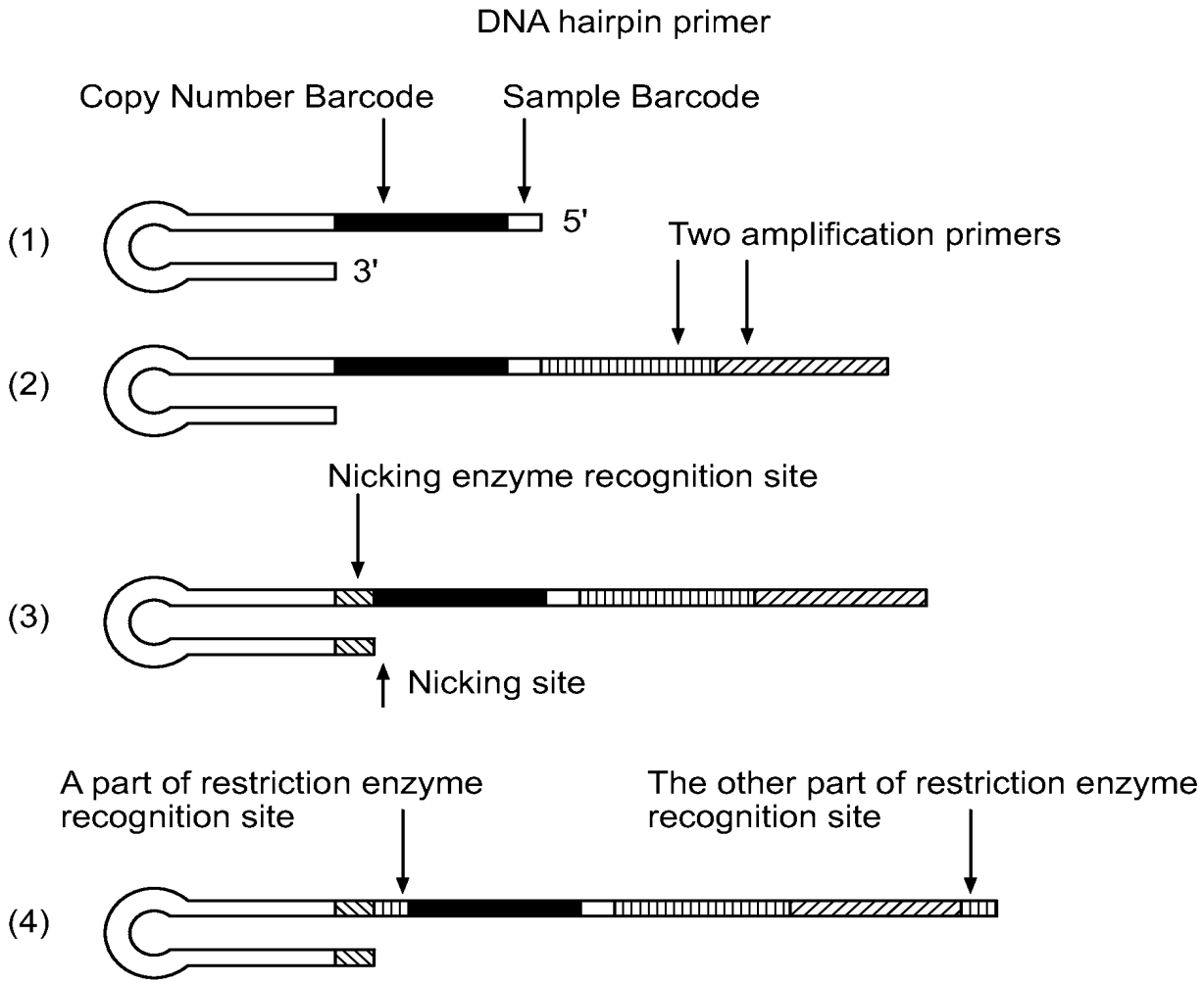

DNA hairpin primer
Copy Number Barcode
Sample Barcode
(1)
5'
3'
Two amplification primers
(2)
Nicking enzyme recognition site
(3)
Nicking site
A part of restriction enzyme recognition site
The other part of restriction enzyme recognition site
(4)

FIG. 2

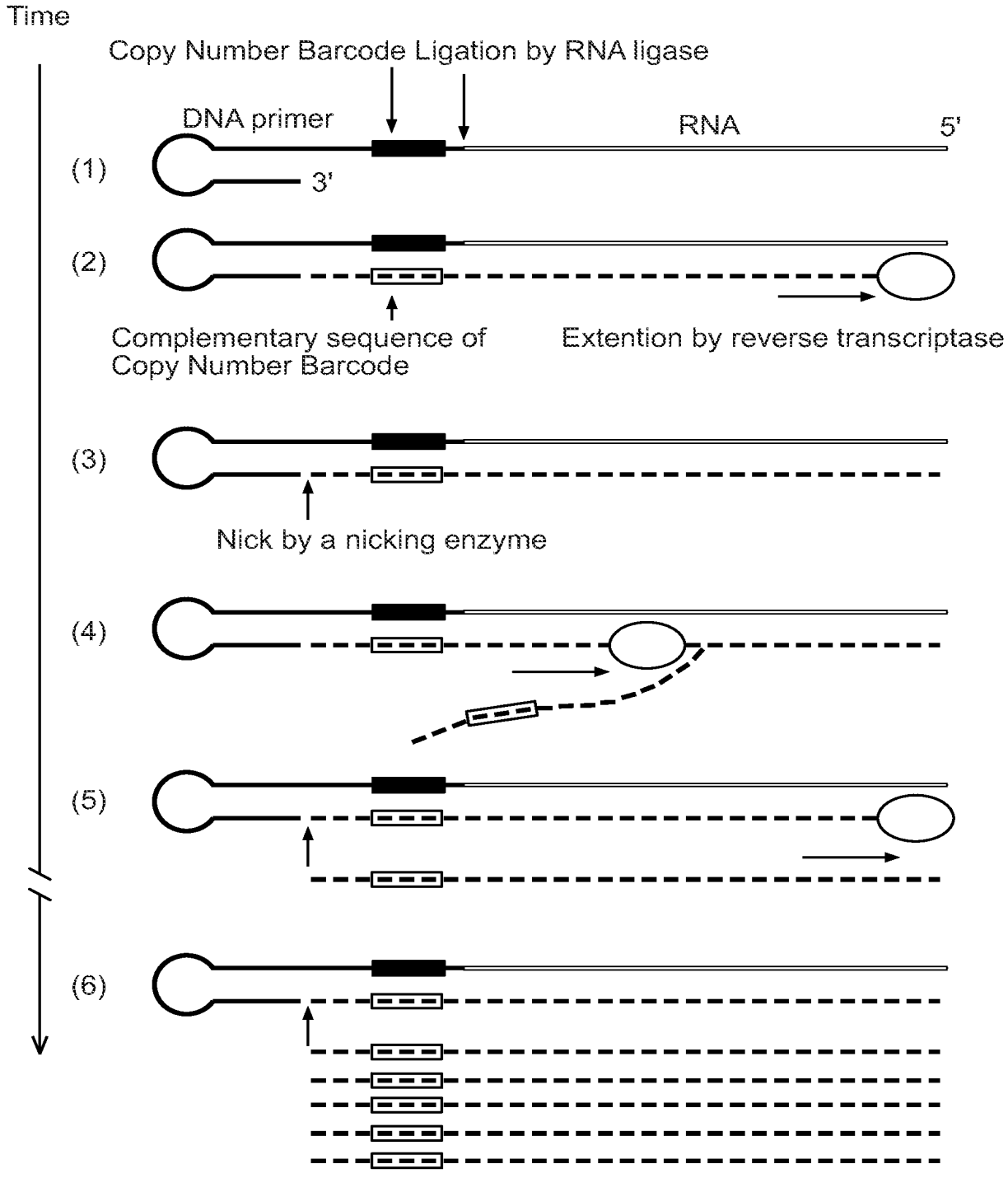

Time
Copy Number Barcode Ligation by RNA ligase
DNA primer
RNA
5'
(1)
3'
(2)
Complementary sequence of
Copy Number Barcode
Extention by reverse transcriptase
(3)
Nick by a nicking enzyme
(4)
(5)
(6)

FIG. 3

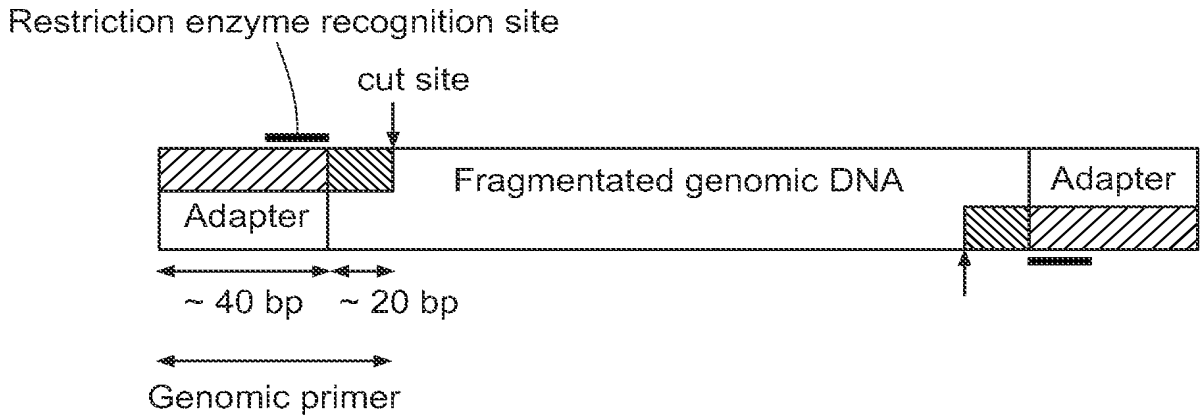


FIG. 4A

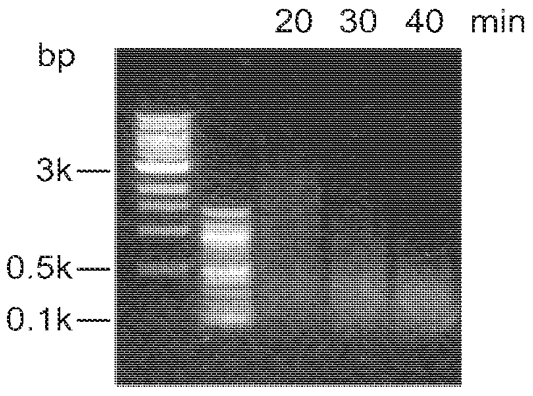


FIG. 4B

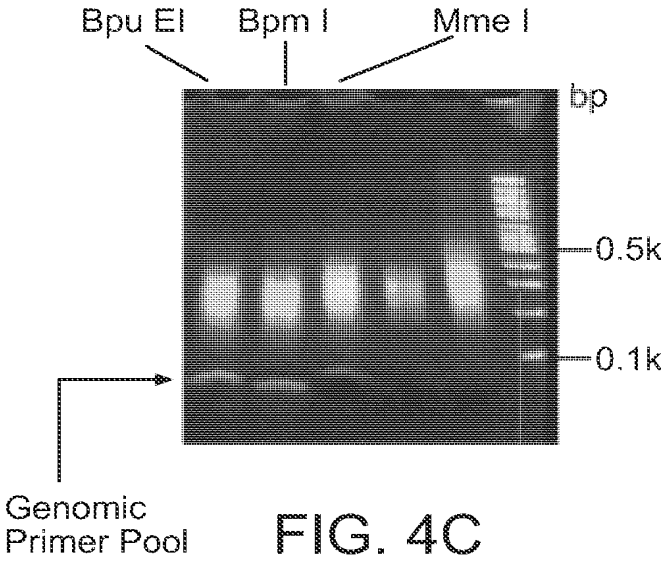


FIG. 4C

FIG. 5

SUBSTITUTE SHEET (RULE 26)

FIG. 6

SUBSTITUTE SHEET (RULE 26)

FIG. 7

SUBSTITUTE SHEET (RULE 26)

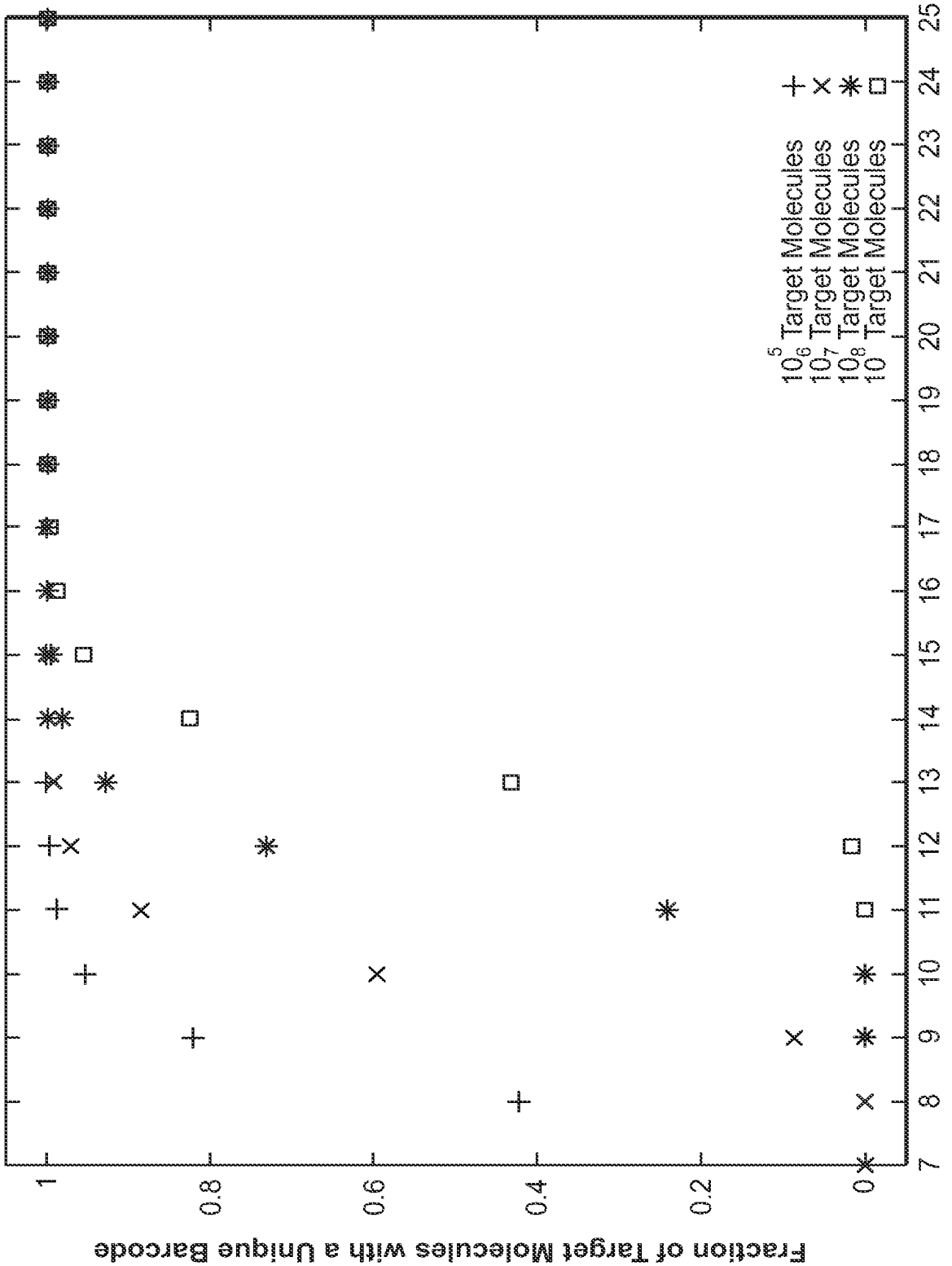

Fraction of Target Molecules with a Unique Barcode
0
0.2
0.4
0.6
0.8
1
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Barcode Length
10^5 Target Molecules
10^6 Target Molecules
10^7 Target Molecules
10^8 Target Molecules

FIG. 8

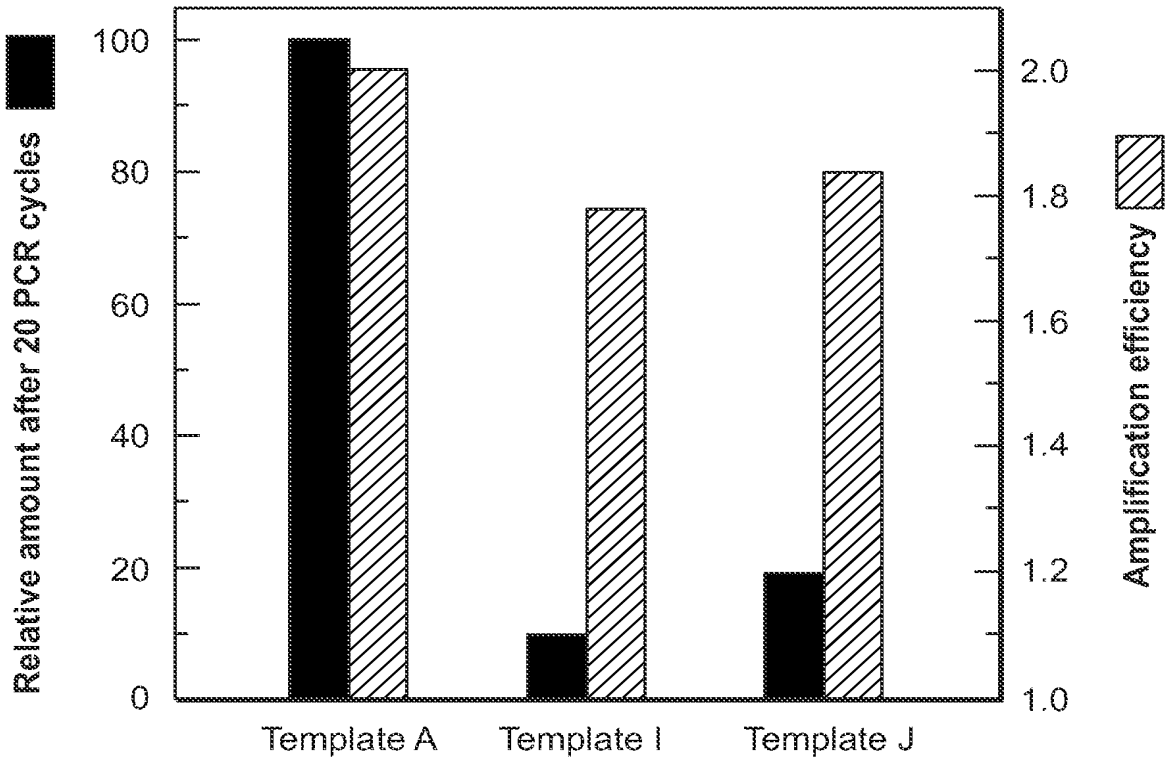

Relative amount after 20 PCR cycles
100
80
60
40
20
0
2.0
1.8
1.6
1.4
1.2
1.0
Amplification efficiency
Template A
Template I
Template J

FIG. 9

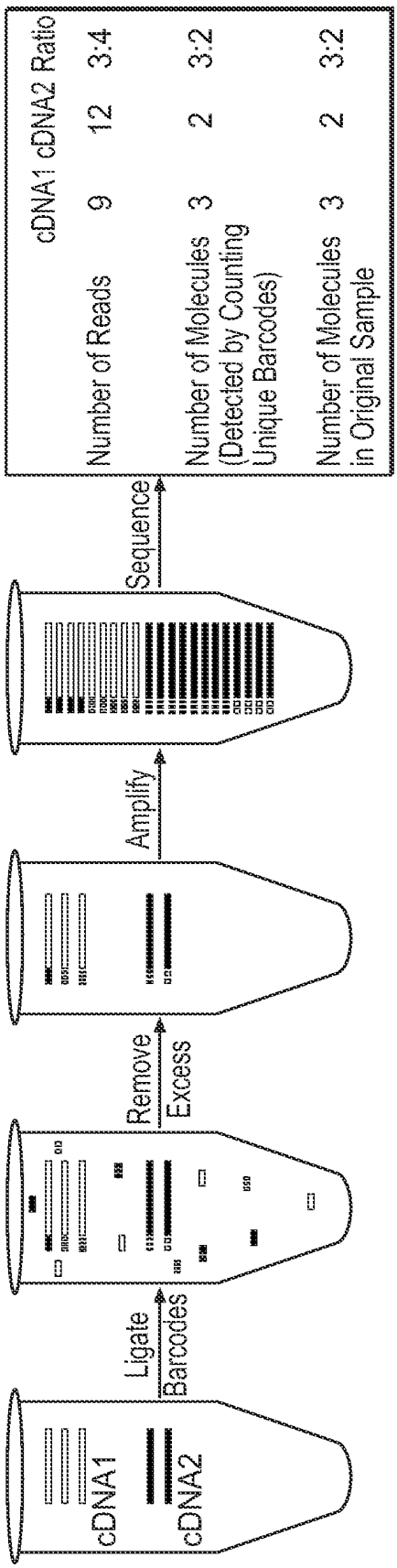


FIG. 10A

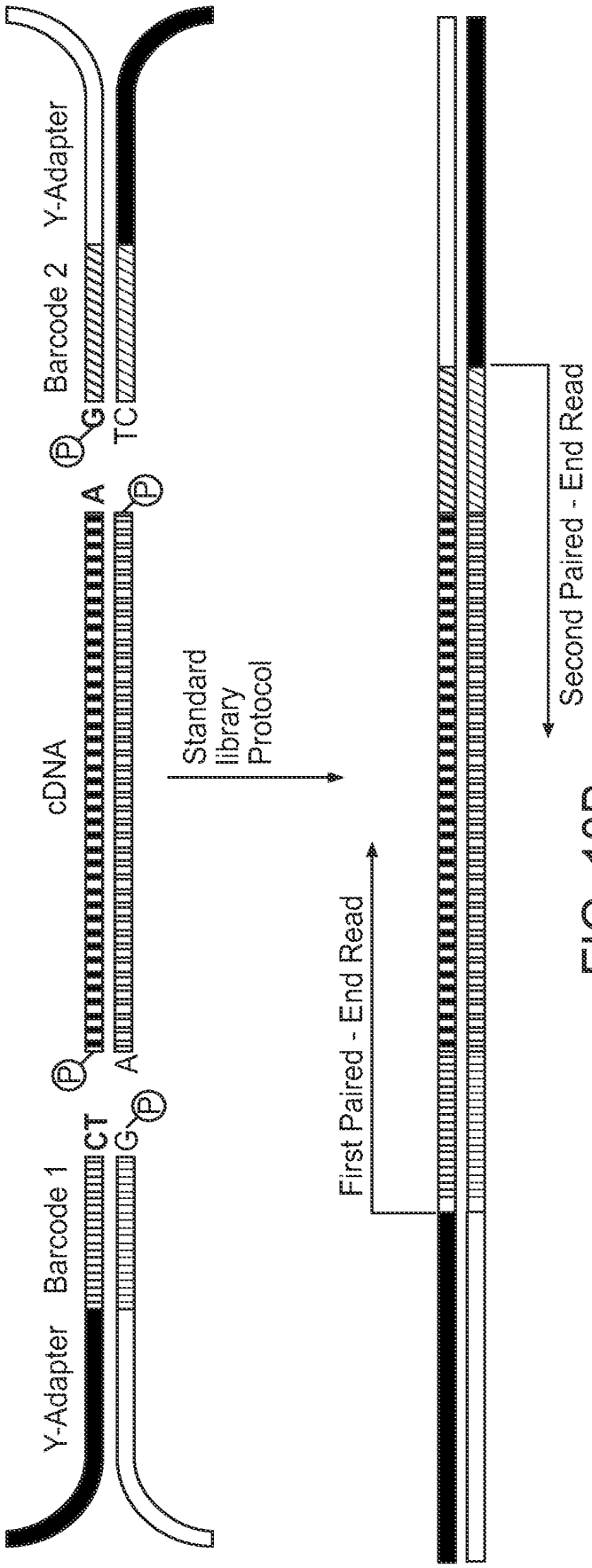


FIG. 10B

FIG. 11A

FIG. 11B

FIG. 11C

SUBSTITUTE SHEET (RULE 26)

FIG. 12A

FIG. 12B

	Number of Barcodes or Barcode Pairs		
	Before Sequencing (Original Solution)	After Sequencing (Simulation)	
		w/o Sequencing Error	w/ Sequencing Error
Random Barcodes	6000	6000	491532
Optimized Barcodes	6000	6000	6000

FIG. 12C

SUBSTITUTE SHEET (RULE 26)

FIG. 13A

FIG. 13B

SUBSTITUTE SHEET (RULE 26)

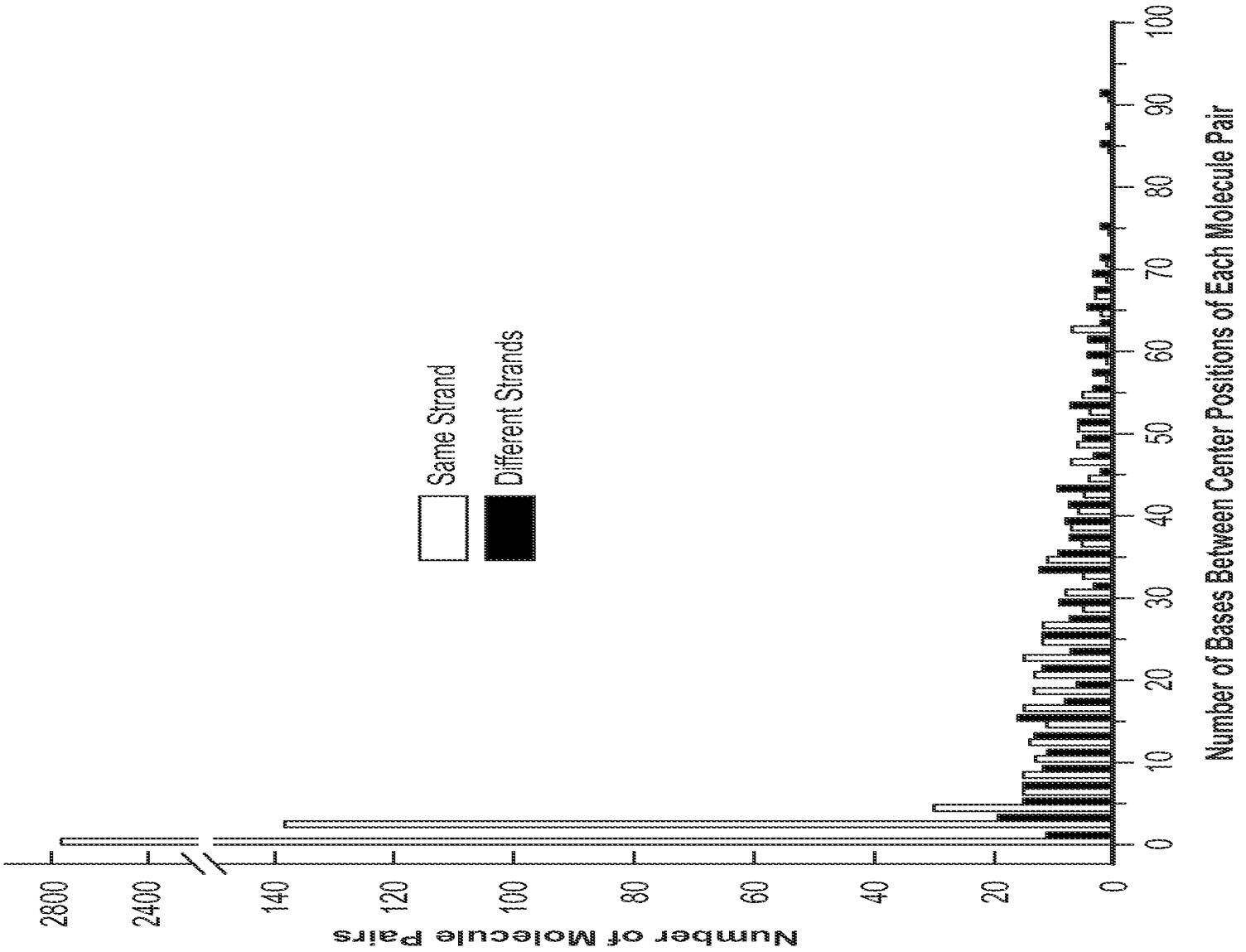

Same Strand
Different Strands
Number of Molecule Pairs
2800
2400
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80
90
100
Number of Bases Between Center Positions of Each Molecule Pair

FIG. 14

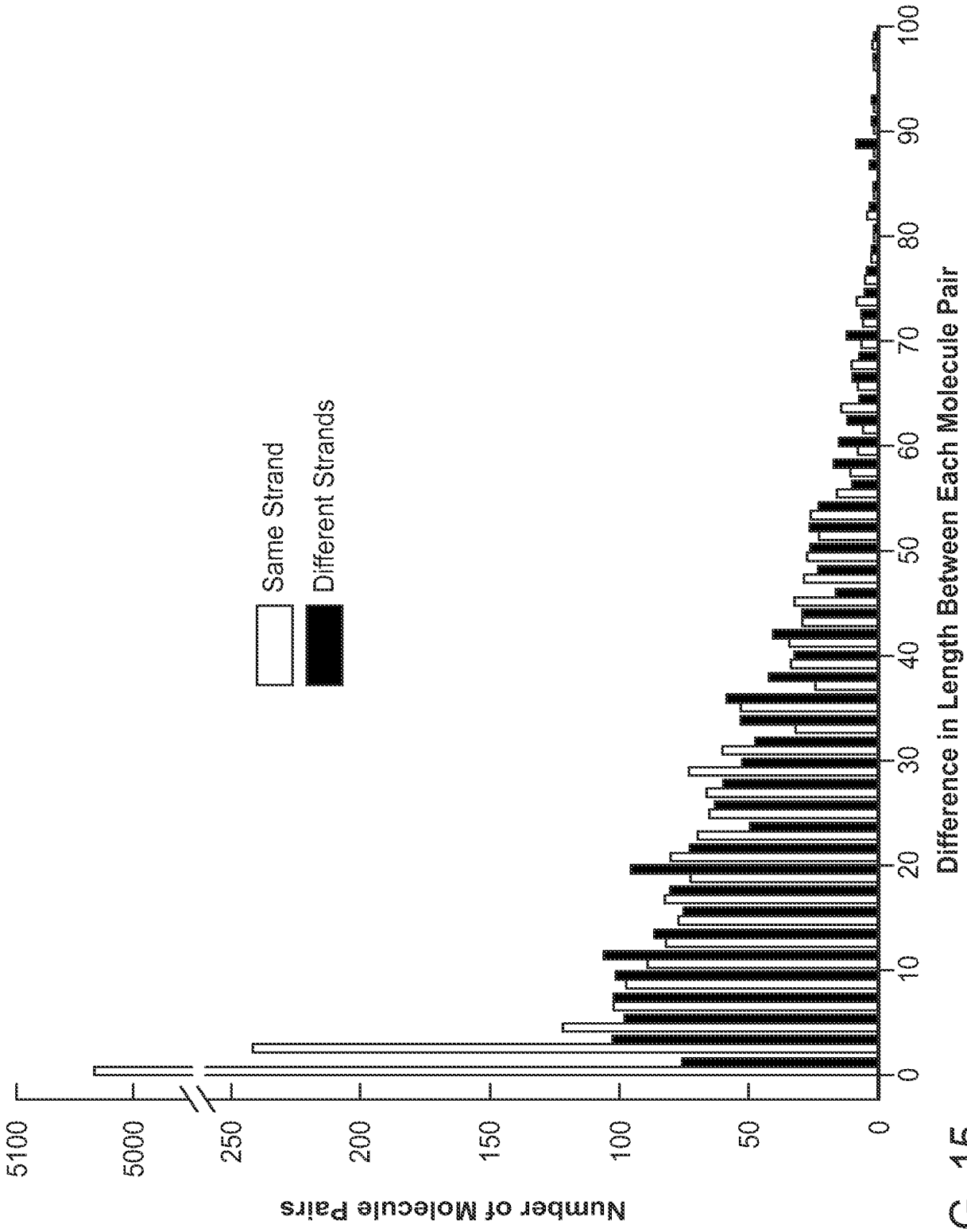

Same Strand
Different Strands
Number of Molecule Pairs
5100
5000
250
200
150
100
50
0
0
10
20
30
40
50
60
70
80
90
100
Difference in Length Between Each Molecule Pair

FIG. 15

FIG. 16A

FIG. 16B

SUBSTITUTE SHEET (RULE 26)

FIG. 16C

FIG. 16D

SUBSTITUTE SHEET (RULE 26)

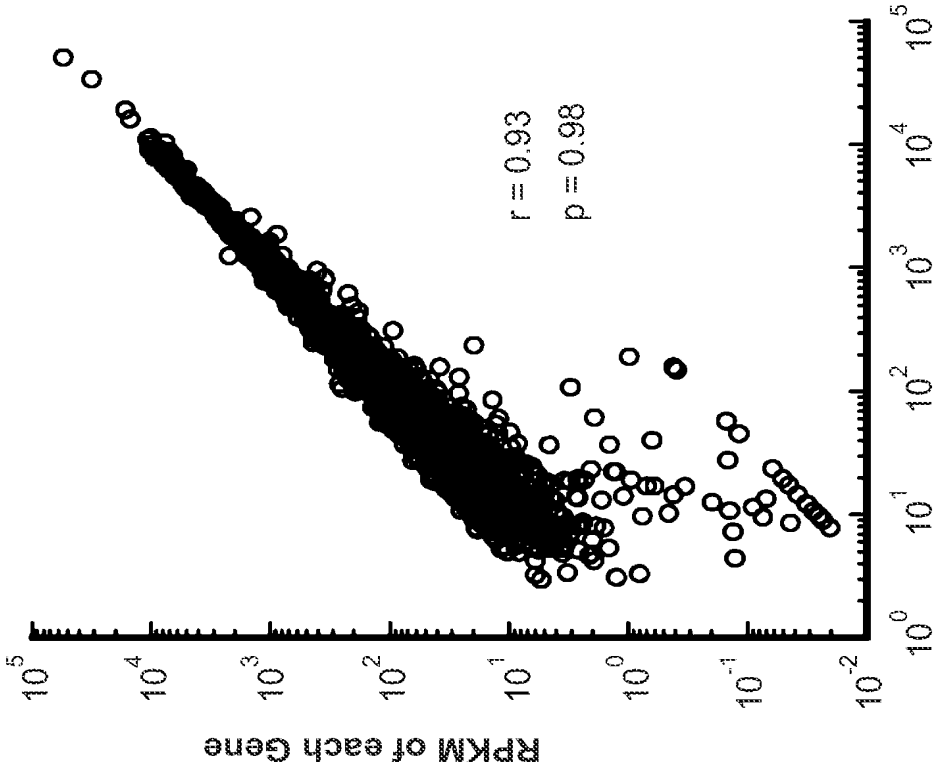


FIG. 17B

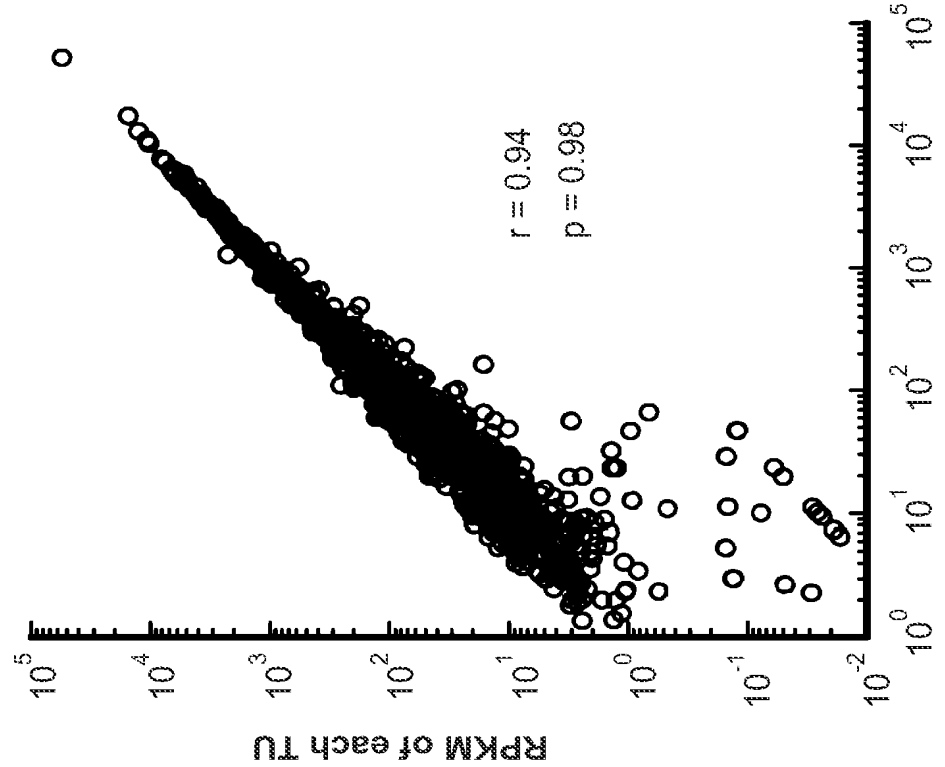


FIG. 17A

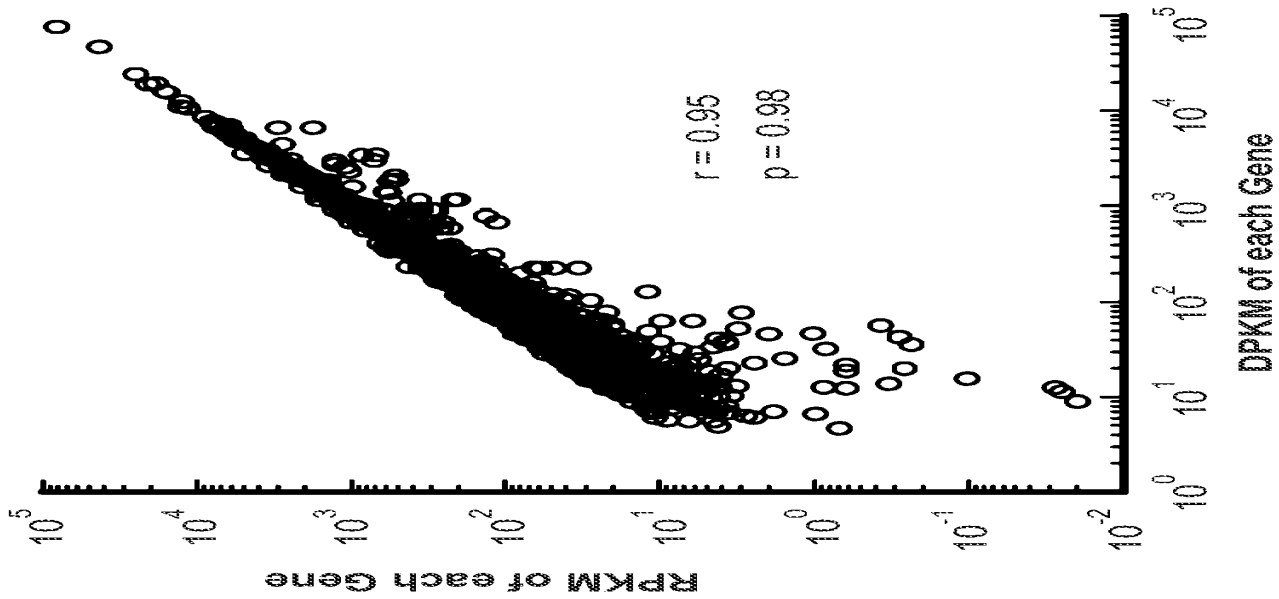


FIG. 17D

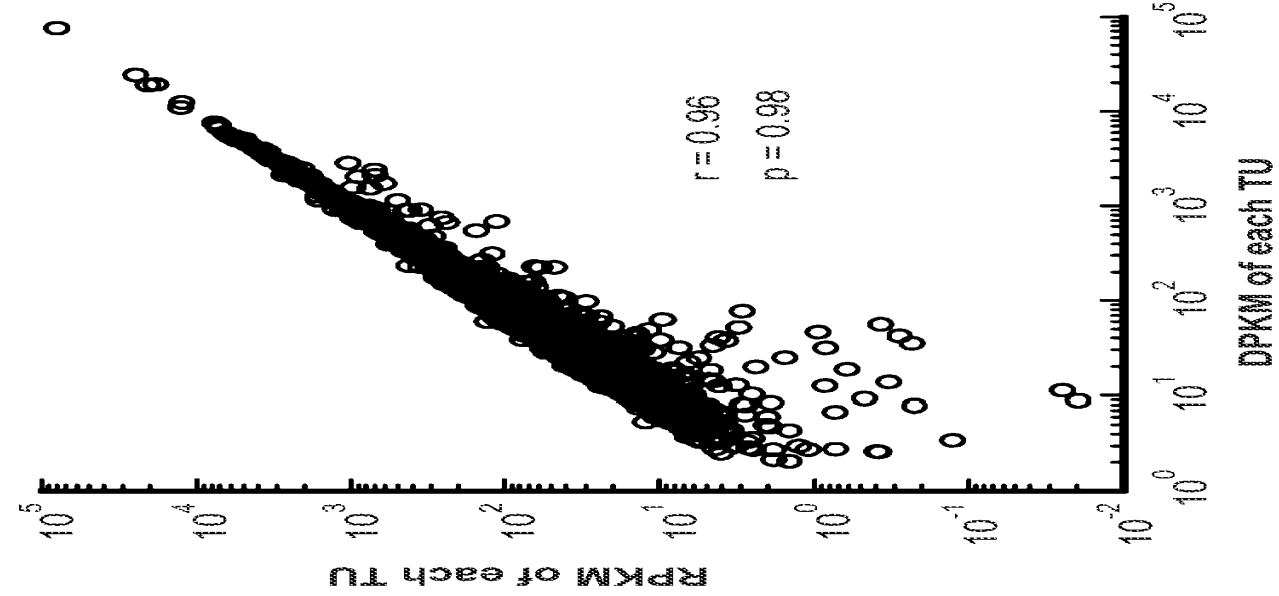


FIG. 17C

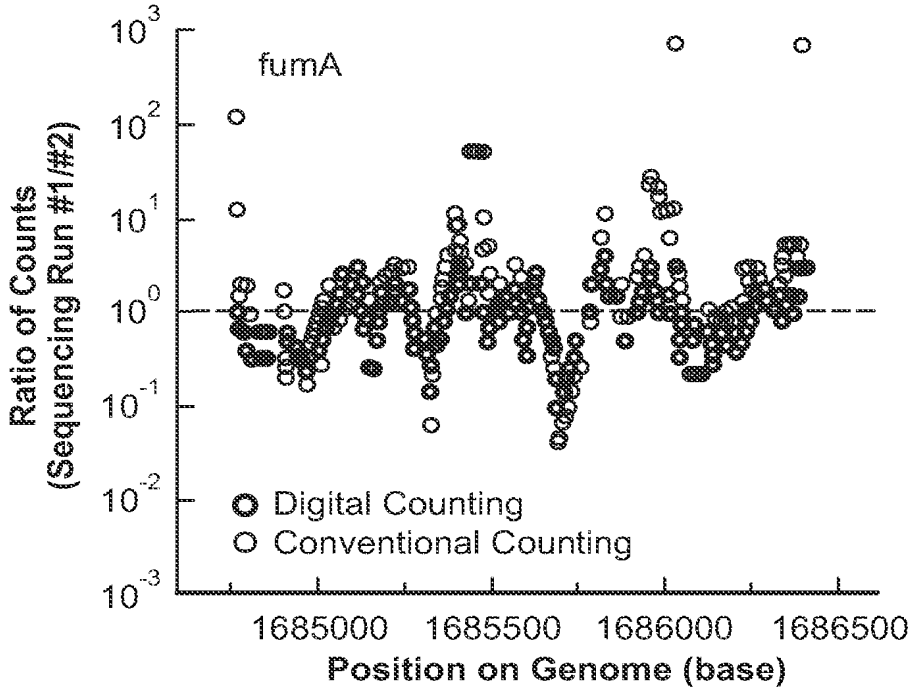


FIG. 18A

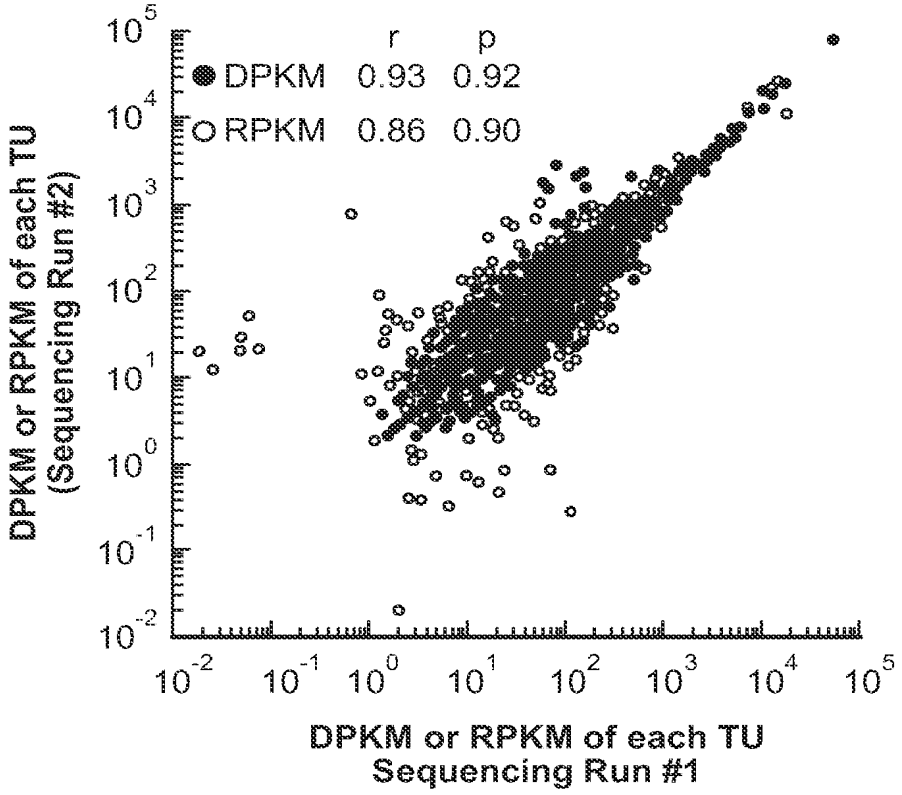


FIG. 18B

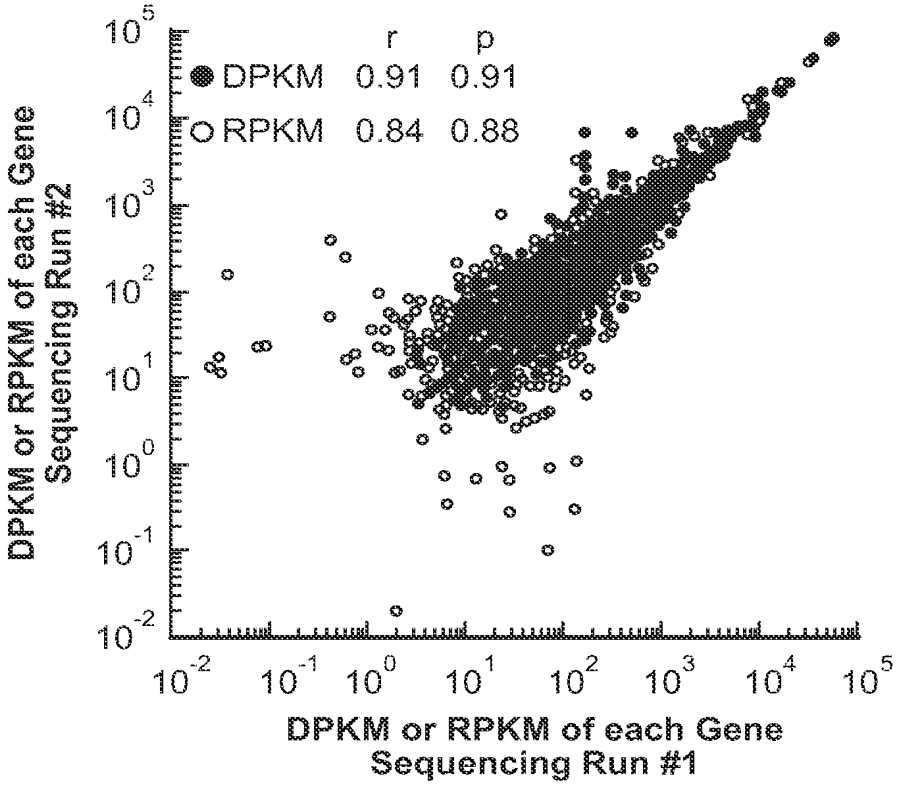


FIG. 18C

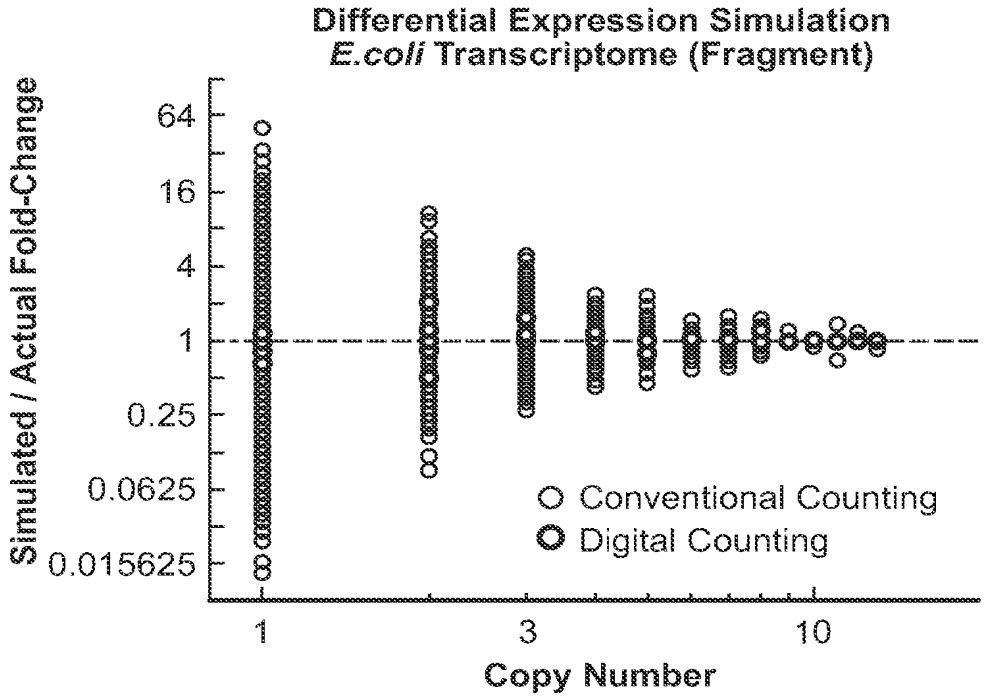


FIG. 19A

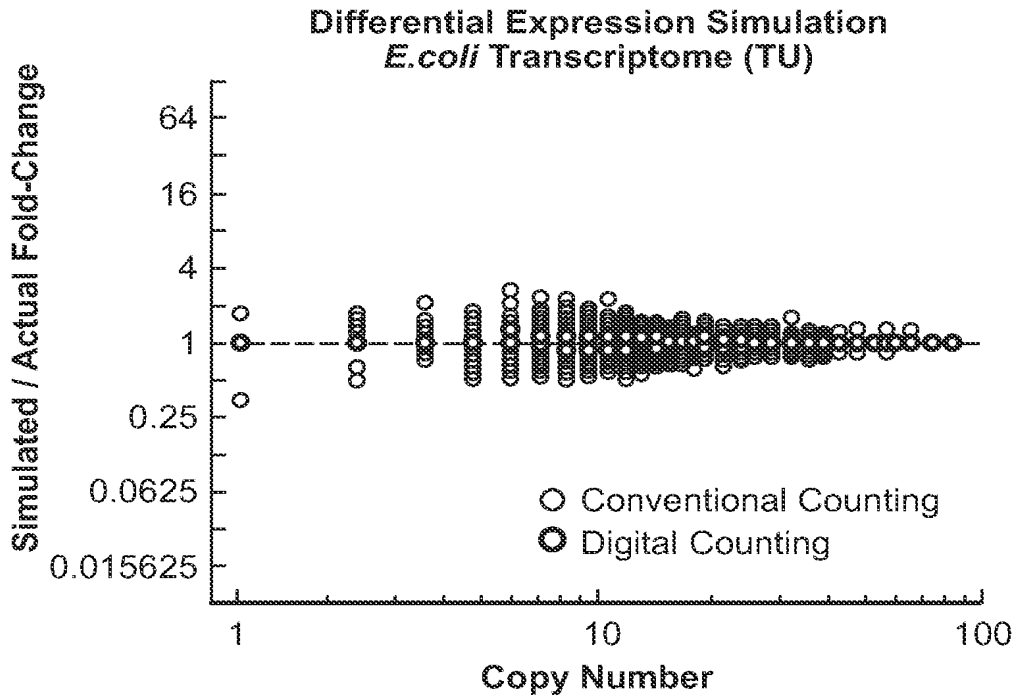


FIG. 19B

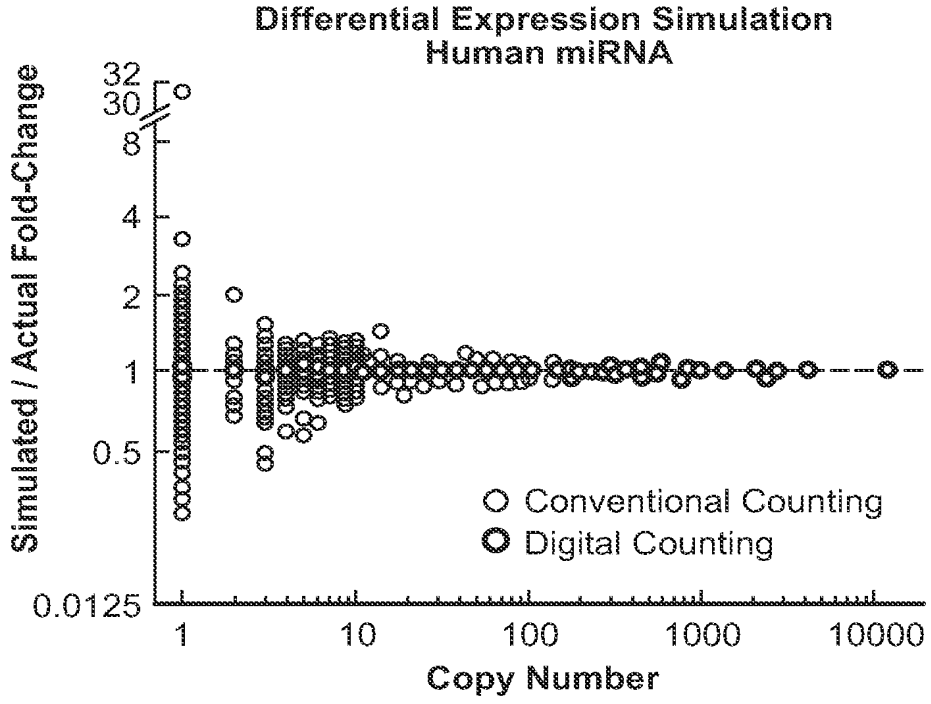


FIG. 19C

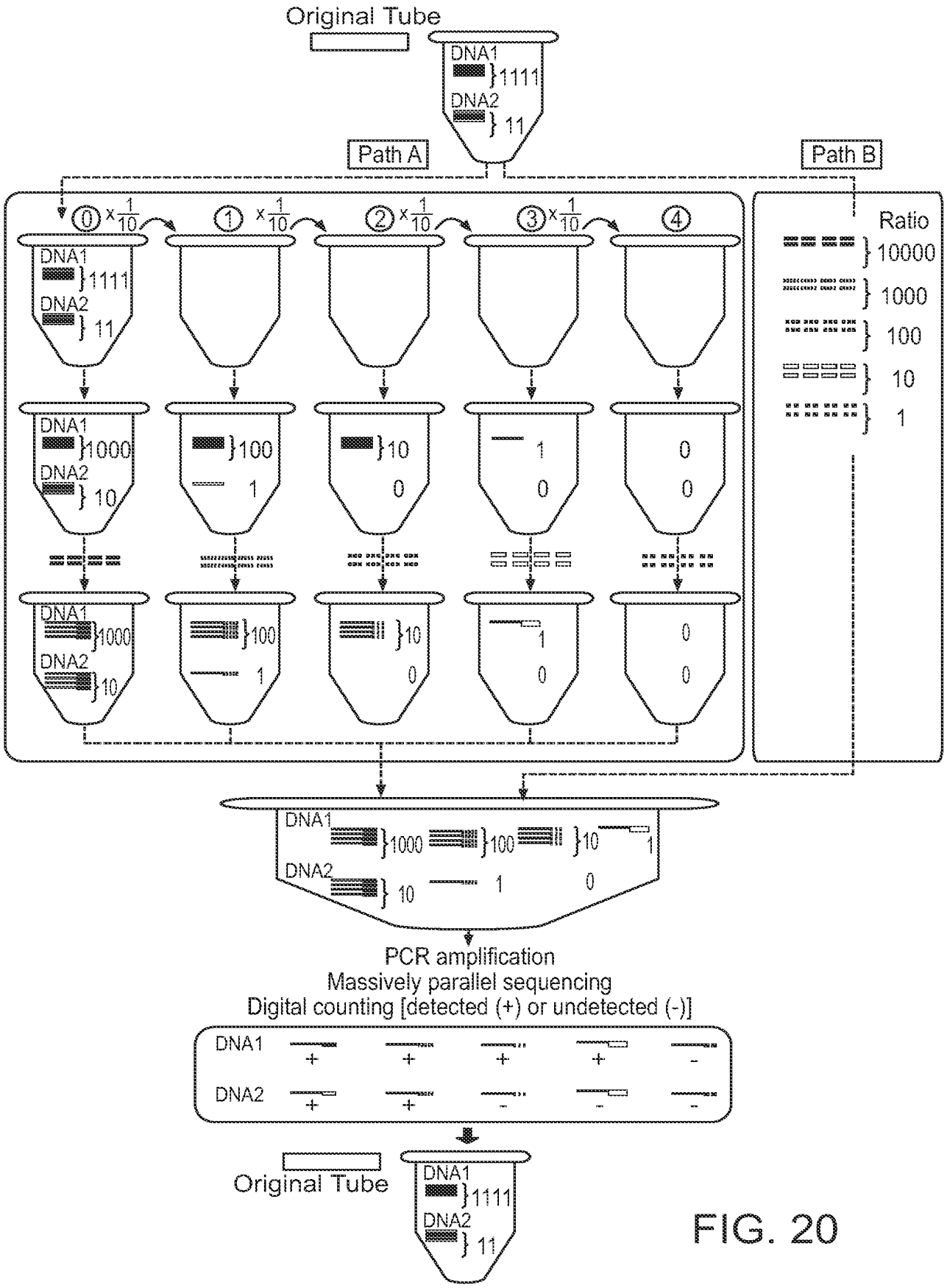

Original Tube
DNA1
1111
DNA2
11
Path A
Path B
0 x 1/10
1 x 1/10
2 x 1/10
3 x 1/10
4
Ratio
10000
1000
100
10
1
DNA1
1000
DNA2
10
100
1
10
0
1
0
0
0
PCR amplification
Massively parallel sequencing
Digital counting [detected (+) or undetected (-)]
DNA1
DNA2
Original Tube

FIG. 20

Electronic Acknowledgement Receipt

EFS ID:	38045968
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	16-DEC-2019
Filing Date:	30-AUG-2018
Time Stamp:	17:58:43
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	Information Disclosure Statement (IDS) Form (SB08)	2019-12-16-Supp-IDS-072227-8043US07.pdf	217953 60bcfba22c05b0d0c4e0cf8b3c5ab80dc528bd25	no	6

Warnings:

Information:					
This is not an USPTO supplied IDS fillable form					
2	Foreign Reference	B1-WO2012042374A2.PDF	1167895 7ad89546db44f295d8749e482f1a5630bad7bc7d	no	65
Warnings:					
Information:					
3	Foreign Reference	B2-WO2012129363A2.PDF	1657351 e8a2a3ee50f9ebbea3f8becf345bfac3d3842fae	no	111
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					
4	Non Patent Literature	C01-Bentley-Nature-456-53-9-2008.PDF	309817 2fc823ccb520f2926077e2432cfa1b461b15491d	no	7
Warnings:					
Information:					
5	Non Patent Literature	C02-Chiu-PNAS-105-10458-63-2008.PDF	898717 66e3f811442adfcad1ac1d80d225eb3f70a5b016	no	6
Warnings:					
Information:					
6	Non Patent Literature	C03-Diehl-Gastroenterology-135-489-98-2008.PDF	781260 aba29fd70ab7ff04ad4a9502ec06655771fdb4b3	no	17
Warnings:					
Information:					
7	Non Patent Literature	C04-Diehl-PNAS-102-16368-73-2005.PDF	421842 e6a762c8c803398d93e680f246594181390309a2	no	6
Warnings:					
Information:					
8	Non Patent Literature	C05-Ding-Hum-Mol-Genetics-19-R188-96-2010.PDF	510471 896c737032ae61c65693512628d5ad42cf6e8305	no	9

Warnings:					
Information:					
9	Non Patent Literature	C06-EP13764186_OA_2017-01-31.PDF	5366574 6a683349030b5400403da39b43ee76e2f326cfaa	no	6
Warnings:					
Information:					
10	Non Patent Literature	C07-Ewing-1998.PDF	607403 63424a838be06f8c960798a87fddb937e28528de	no	10
Warnings:					
Information:					
11	Non Patent Literature	C08-Fleischhacker-Biochim-Biophys-1775-181-232-2007-2.PDF	4723333 1768aed773e10ea4f29ff8f78a47d690c3288ce3	no	52
Warnings:					
Information:					
12	Non Patent Literature	C09-Fong-Clin-Chem-55-587-98-2009.PDF	630041 9a46f46e842ddc281f92418fe1375568d404c3ff	no	12
Warnings:					
Information:					
13	Non Patent Literature	C10-Forshew-2012.PDF	667885 db0a1b5f66ce96062783ccdd2596dc7ae88ed87c	no	14
Warnings:					
Information:					
14	Non Patent Literature	C11-Gordon-Nat-Rev-Genet-13-189-203-2012-2.PDF	2381826 776117bed9db52003e2161c13bbbce48434c8124	no	15
Warnings:					
Information:					
15	Non Patent Literature	C12-Jiang-Bioinformat-24-2395-6-2008.PDF	62467 987a44632a78ae31afee4c6e9b635d4bded5d468	no	2
Warnings:					
Information:					

16	Non Patent Literature	C13-Kao-Genome-Res-19-1884-95-2009.PDF	1184438 080c2e13fa70b94fe4d35799b7c883cf69164a96	no	13
Warnings:					
Information:					
17	Non Patent Literature	C14-Kircher-Genome-Biol-10-R83-2009-2.PDF	241155 26c0ac5e690023cdc4dcbc19fcff6c532f484b6e	no	9
Warnings:					
Information:					
18	Non Patent Literature	C15-Kirsch-PNAS-109-14289-90-2012.PDF	201395 c9948fb76827fb0d5e3711983a41cd83a8748cea	no	2
Warnings:					
Information:					
19	Non Patent Literature	C16-Ledergerber-Brief-Bioinformat-12-489-97-2011.PDF	325945 db19a5c04b2f12acb0bb5d7814225f7f397d3257	no	9
Warnings:					
Information:					
20	Non Patent Literature	C17-Li-Bioinformat-25-1754-60-2009.PDF	154338 b7cdba802946ccb11b611d0c96ec9105cec20dd8	no	7
Warnings:					
Information:					
21	Non Patent Literature	C18-Li-Genome-Res-18-1851-8-2008.PDF	566242 39ccb506cb6bdac1fbc812885aa53d3222cf7cbe	no	9
Warnings:					
Information:					
22	Non Patent Literature	C19-Liang-Trans-Comp-Biol-Bioinformat-4-430-40-2007.PDF	687660 f16ba22a27d078746cfdd539ecff73c811e93ace	no	11
Warnings:					
Information:					

23	Non Patent Literature	C20-Liao-Clin-Chem-57-92-101-2011.PDF	1027616 71c1739b28c8151147906aeca41de17e52188aea	no	10
Warnings:					
Information:					
24	Non Patent Literature	C21-Lo-Am-J-Hum-Genet-62-768-75-1998.PDF	648392 3509f93d0343263cc90cc920a7a6d4965fc90691	no	8
Warnings:					
Information:					
25	Non Patent Literature	C22-Lunter-Genome-Res-21-936-9-2011.PDF	640171 fb5211157a73e961b8a27f7150fc233dc196722d	no	6
Warnings:					
Information:					
26	Non Patent Literature	C23-McKernan-Genome-Res-19-1527-41-2009.PDF	1390171 d17ce1f6056ec6f0794bf189c21e59f84451eebf	no	16
Warnings:					
Information:					
27	Non Patent Literature	C24-Mertes-Brief-Func-Genomics-10-374-86-2011.PDF	1069359 6e8cba4f31cb1a2c847e68c020a5ab735a4a769b	no	13
Warnings:					
Information:					
28	Non Patent Literature	C25-Meyer-2007-Article.PDF	639167 b55e557f560ba3ad92fd5b1284ad9151ef1cef92	no	5
Warnings:					
Information:					
29	Non Patent Literature	C26-Meyerson-Nat-Rev-Genetics-11-685-96-2010-2.PDF	1005964 0bcba32a70eb8a5ba9190264d74aba0db124565d	no	12
Warnings:					
Information:					

30	Non Patent Literature	C27-Narayan-2012.PDF	349777 d141a06254c66d346374afa48bb99db97f844874	no	8
Warnings:					
Information:					
31	Non Patent Literature	C28-Nielsen-Nat-Rev-Genetics-12-443-51-2011-2.PDF	537648 ef44ced14d011b30e66a5e0f7a0ae040a5c7390d	no	9
Warnings:					
Information:					
32	Non Patent Literature	C29-Quinlan-Nat-Methods-5-179-81-2008.PDF	226126 d43e47dcf0c6df9a55bcd3b5d08ee07076e72b13	no	3
Warnings:					
Information:					
33	Non Patent Literature	C30-Redon-Nature-444-444-54-2006.PDF	1086760 86a060ed1f80df98e02c51aaec2bcf12eeaedc8c	no	11
Warnings:					
Information:					
34	Non Patent Literature	C31-Rizzo-Buck-2012.PDF	771705 db18ec6788e055cdfae018c35aeab2369b8b2978	no	15
Warnings:					
Information:					
35	Non Patent Literature	C32-Schmitt-PNAS-109-14508-13-2012.PDF	237445 fbabad94fbe322104595864eada70faa4109b44a	no	6
Warnings:					
Information:					
36	Non Patent Literature	C33-Schwarzenbach_Nature_Reviews_11_426-37_2011.PDF	417045 020f0ba13c0b8f2255b1efa59ca12ce8db7b9e65	no	12
Warnings:					
Information:					

37	Non Patent Literature	C34-Schweiger-PLOS-4-e5548-2009.PDF	554616 0b72da272c6bd0ae4a800ae380482a8bf31bc195	no	7
Warnings:					
Information:					
38	Non Patent Literature	C35-Sehnert-Clin-Chem-57-1042-49-2011.PDF	283963 5023fa9ec1f1f560ce3f637e315112d306fb28ca	no	8
Warnings:					
Information:					
39	Non Patent Literature	C36-Sparks-Prenatal-Diag-32-3-9-2012.PDF	248629 5ba7a441f671f381ab90b2a582c84f8fa2d57abf	no	7
Warnings:					
Information:					
40	Non Patent Literature	C37-Teer-Genome-Res-20-1420-31-2010.PDF	955141 3d9de9db17e816273a108506f90bfc4a45034db1	no	13
Warnings:					
Information:					
41	Non Patent Literature	C38-US16120019_NOA_2019-10-30.PDF	797492 8bf91fa216530326e55b393aee8ab937beca58bc	no	15
Warnings:					
Information:					
42	Non Patent Literature	C39-US16120072_NOA_2019-06-26.PDF	623640 9a27cb529495938e2efd0d8e86cb69c29b057ded	no	11
Warnings:					
Information:					
43	Non Patent Literature	C40-US16120091_NOA_2019-06-10.PDF	1886999 639b0ba6cbeb8a5bd42ace3a8c68707ff4696402	no	8
Warnings:					
Information:					

44	Non Patent Literature	C41-US16410045_NFOA_2019-07-25.PDF	3962815 b34f2b1defc448605d2eb2990de9c887ec8726ae	no	20
Warnings:					
Information:					
45	Non Patent Literature	C42-US16411045_NOA_2019-11-15.PDF	2271131 3675dd7bb67ee3769d57e59ceb4655ae2955d88b	no	16
Warnings:					
Information:					
46	Non Patent Literature	C43-US16411068_NFOA_2019-11-20.PDF	3657524 99544231f04eb41e4dd0e5c19712bb0e162fa72f	no	28
Warnings:					
Information:					
47	Non Patent Literature	C44-US16411069_NFOA_2019-11-26.PDF	3744466 22078c96b133bfebfce898f00f084c2feb0cbbe8	no	28
Warnings:					
Information:					
48	Non Patent Literature	C45-Wagle-Cancer-Disc-2-82-93-2011.PDF	4760913 582b62addc24da38c697fd4ce435a6d8497e96d1	no	13
Warnings:					
Information:					
			Total Files Size (in bytes):	57560683	

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.

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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

94991 7590 08/28/2020

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
08/28/2020	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):

patentprocurement@perkinscoie.com

Office Action Summary	Application No. 16/118,306	Applicant(s) Salk et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status No

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

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 2 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

☐ 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) 37-66 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☐ Claim(s) _____ is/are rejected.

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

9) ☑ Claim(s) 37-66 are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions.

Species election

2. This application contains claims directed to the following patentably distinct species:

Species of variations present in the consensus sequence are indicative of: (claim 51)

A) a cancer risk

B) a cancer metabolic state

C) a mutator phenotype

D) a carcinogen exposure

E) a chronic inflammation exposure

F) an age

G) a neurodegenerative disease

Species of location of molecule-specific feature: (claim 53)

H) molecule-specific feature is at an end of the dsDNA molecule

I) molecule-specific feature is adjacent to an end of the dsDNA molecule

J) molecule-specific feature is at a specified position relative to an end of the dsDNA molecule

Species of adapter molecule: (claim 57)

K) comprises a single-stranded SMI sequence

L) comprises a double-stranded SMI sequence

Species of adapter molecule: (claim 59)

M) comprises a Y-shape

N) comprises a U-shape

Species of set of duplicates of original forward and complement strands: (claims 63 and 64)

O) includes a minimum of 1 member in each set

P) includes a minimum of 3 members in each set

Species of source of duplex DNA molecules in a subject: (claim 65)

Q) derived from a tumor

R) derived from circulating neoplastic cells

3. The species are independent or distinct because claims to the different species recite the mutually exclusive characteristics of such species. In addition, these species are not obvious variants of each other based on the current record.

Applicant is required under 35 U.S.C. 121 to elect a single disclosed species from each of the sets of species for prosecution on the merits to which the claims shall

be restricted if no generic claim is finally held to be allowable. Currently, claims 46, 52, 55 and 56 are generic.

There is an examination and search burden for these patentably distinct species due to their mutually exclusive characteristics. The species require a different field of search (e.g., searching different classes/subclasses or electronic resources, or employing different search queries); and/or the prior art applicable to one species would not likely be applicable to another species; and/or the species are likely to raise different non-prior art issues under 35 U.S.C. 101 and/or 35 U.S.C. 112, first paragraph.

4. **Applicant is advised that the reply to this requirement to be complete must include (i) an election of a species to be examined** even though the requirement may be traversed (37 CFR 1.143) **and (ii) identification of the claims encompassing the elected species,** including any claims subsequently added. An argument that a claim is allowable or that all claims are generic is considered nonresponsive unless accompanied by an election.

The election of the species may be made with or without traverse. To preserve a right to petition, the election must be made with traverse. If the reply does not distinctly and specifically point out supposed errors in the election of species requirement, the election shall be treated as an election without traverse. Traversal must be presented at the time of election in order to be considered timely. Failure to timely traverse the requirement will result in the loss of right to petition under 37 CFR 1.144. If claims are added after the election, applicant must indicate which of these claims are readable on the elected species.

Should applicant traverse on the ground that the species are not patentably distinct, applicant should submit evidence or identify such evidence now of record showing the species to be obvious variants or clearly admit on the record that this is the case. In either instance, if the examiner finds one of the species unpatentable over the prior art, the evidence or admission may be used in a rejection under 35 U.S.C. 103(a) of the other species.

Upon the allowance of a generic claim, applicant will be entitled to consideration of claims to additional species which depend from or otherwise require all the limitations of an allowable generic claim as provided by 37 CFR 1.141.

5. Applicant is reminded that upon the cancellation of claims to a non-elected invention, the inventorship must be amended in compliance with 37 CFR 1.48(b) if one or more of the currently named inventors is no longer an inventor of at least one claim remaining in the application. Any amendment of inventorship must be accompanied by a request under 37 CFR 1.48(b) and by the fee required under 37 CFR 1.17(i).

Correspondence

6. Any inquiry concerning this communication or earlier communications from the examiner should be directed to David C. Thomas whose telephone number is 571-272-3320 and whose fax number is 571-273-3320. The examiner can normally be reached on 5 days, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on 571-272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/David C Thomas/
Primary Examiner, Art Unit 1637

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	JESSE SALK, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1637
FILED:	AUGUST 30, 2018	EXAMINER:	DAVID C. THOMAS
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

RESPONSE TO RESTRICTION REQUIREMENT AND AMENDMENT

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

Dear Commissioner:

In response to the restriction requirement mailed August 28, 2020, please enter the elections on page 10 of this paper, and amend the application as follows:

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

Remarks begin on page 10 of this paper.

IN THE CLAIMS

The following complete listing of claims replaces all previous listings of claims in the application.

1-36. (Canceled)

37. (Previously Presented) A method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising:

amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products;

sequencing the amplified target nucleic acid products generated from each original strand;

confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands;

comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand; and

generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary.

38. (Previously Presented) The method of claim 37, wherein prior to amplifying each strand, the method further comprises ligating the double-stranded target nucleic acid molecule to an adapter molecule resulting in an asymmetric adapter-target nucleic acid complex such that each original strand generates a distinct yet related set of amplified target nucleic acid products during the amplification step.

39. (Previously Presented) The method of claim 38, wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

40. (Previously Presented) The method of claim 39, wherein the adapter molecule further comprises a nucleotide sequence that uniquely tags each original strand of the double-stranded target nucleic acid molecule such that each original strand is distinctly identifiable relative to its original complementary strand.

41. (Previously Presented) The method of claim 37, wherein the confirming step includes (i) grouping the sequence reads of the amplified target nucleic acid products into families of paired target nucleic acid strand sequences using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the double-stranded target nucleic acid molecule as an identifier, and (ii) separating each individual family of paired target nucleic acid strand sequences into a set of sequence reads from one original strand of the double-stranded target nucleic acid molecule and a set of sequence reads from the other original strand of the double stranded target nucleic acid molecule.

42. (Previously Presented) The method of claim 37, wherein the confirming step includes confirming the presence of at least about 3 sequence reads of an amplified target nucleic acid product generated from each of the original strands, and wherein the comparing step includes comparing a first consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from one original strand with a second consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from the other original strand.

43. (Previously Presented) The method of claim 37, further comprising identifying a mutation occurring at a particular position in the consensus sequence as a true mutation.

44. (Previously Presented) The method of claim 43, further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation.

45. (Previously Presented) The method of claim 37, wherein generating a consensus sequence of the double-stranded target nucleic acid molecule further comprises identifying one or more nucleotide positions that disagree between the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand, and discounting nucleotide positions in the consensus sequence that are in disagreement to provide an error-corrected sequence read of the double-stranded target nucleic acid molecule.

46. (Previously Presented) A method for generating a high accuracy DNA sequencing read, comprising:

sequencing a first strand of a double-stranded DNA (dsDNA) molecule to generate a first strand sequencing read;

sequencing a second strand of the same dsDNA molecule to generate a second strand sequencing read;

grouping the first and second sequence reads by a common feature;

comparing the first strand sequencing read to the second strand sequencing; and

generating a consensus sequence comprising a sequence of nucleotide bases, wherein each nucleotide base is identified at a given position in the consensus

sequence when a specific nucleotide is complementary between the first strand sequencing read and the second strand sequencing read.

47. (Previously Presented) The method of claim 46, wherein generating a consensus sequence comprises:

identifying nucleotide positions where the compared first and second strand sequence reads are non-complementary;

identifying nucleotide positions where the compared first and second strand sequence reads are complementary; and

generating a high accuracy consensus sequence read for the dsDNA molecule that includes only the nucleotide positions where the compared first and second strand sequence reads are complementary.

48. (Previously Presented) The method of claim 46, further comprising comparing the consensus sequence to a respective reference sequence, and identifying one or more nucleotide sequence variations present in the consensus sequence not present in the reference sequence.

49. (Previously Presented) The method of claim 48, further comprising identifying a variation occurring at a particular position in the consensus sequence as a true mutation.

50. (Currently Amended) The method of claim [[48]]46, further comprising identifying a potential artifact ~~that occurs~~ at a particular position when a base call in ~~only one of~~ the first strand sequence read ~~or~~ does not agree with the base call in the second strand sequence read ~~as a potential artifact~~.

51. (Previously Presented) The method of claim 46, wherein the dsDNA molecule is one of a population of dsDNA molecules, and wherein the method further

comprises generating a consensus sequence for a plurality of the dsDNA molecules, and wherein for each of one or more of the generated consensus sequences, the method further comprises:

comparing the consensus sequence to a reference sequence; and

identifying one or more variations present in the consensus sequence not present in the reference sequence,

wherein the one or more variations present in the consensus sequence are indicative of a cancer risk, a cancer metabolic state, a mutator phenotype, a carcinogen exposure, a chronic inflammation exposure, an age, a neurodegenerative disease, or a combination thereof in an organism from which the dsDNA molecule is derived.

52. (Previously Presented) The method of claim 46, wherein the common feature comprises one or more DNA molecule features that distinguish the dsDNA molecule from other dsDNA molecules in a sample.

53. (Previously Presented) The method of claim 52, wherein the one or more DNA molecule features includes a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the dsDNA molecule, or a combination thereof.

54. (Currently Amended) The method of claim 46, wherein prior to sequencing either the first strand or the second strand, the method further comprises ligating the dsDNA molecule to at least one adapter molecule to form an adapter-dsDNA complex, and wherein the adapter molecule comprises a single molecule identifier (SMI) sequence that alone or in combination with ~~fragment~~ one or more DNA molecule features of the dsDNA molecule, provides the common feature.

55. (Currently Amended) A method of generating an error-corrected consensus sequence comprising:

ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules;

for each of one or more tagged DNA molecules[[,]]

generating a set of duplicates of an original forward strand of the tagged DNA molecule and a set of duplicates of an original complement strand of the tagged DNA molecule;

creating a single strand consensus sequence (SSCS) from the set of duplicates of the original forward strand and a single strand consensus sequence (SSCS) from the set of duplicates of the original complement strand;

comparing the SSCS of the original forward strand to the SSCS of the original complement strand; and

generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward strand and the SSCS of the original complement strand are complimentary.

56. (Currently Amended) The method of claim 55, wherein [[the]] one or more adapter molecules are[[is a]] single molecule identifier (SMI) adapter molecules comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

57. (Currently Amended) The method of claim 56, wherein the SMI adapter molecules comprise[[s]] a single-stranded SMI sequence or a double-stranded SMI sequence.

58. (Previously Presented) The method of claim 56, wherein the tagged DNA molecule comprises an SMI sequence in each of its original forward strand and original complement strand, and preferably wherein the SMI sequence in one strand comprises

a first degenerate or semi-degenerate sequence and in the other strand comprises a second degenerate or semi-degenerate sequence that is complementary to the first such sequence.

59. (Currently Amended) The method of claim [[56]]55, wherein the adapter molecules comprises a Y-shape, a U-shape, or a combination thereof.

60. (Previously Presented) The method of claim 55, wherein the tagged DNA molecule comprises an adapter molecule at each end.

61. (Currently Amended) The method of claim 55, wherein the comparing step includes pairing the SSCS of the original forward strand to the SSCS of the original complement strand using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of ~~the~~ a particular duplex DNA molecule[[s]] as an identifier.

62. (Previously Presented) The method of claim 55, further comprising sequencing one or more of the duplicates of the original forward strand and one or more of the duplicates of the original complement strand of the tagged DNA molecules using a next-generation sequencing procedure.

63. (Previously Presented) The method of claim 55, wherein the set of duplicates of the original forward strand and the set of duplicates of the original complement strand includes a minimum of 1 member in each set.

64. (Previously Presented) The method of claim 55, wherein the set of duplicates of the original forward strand and the set of duplicates of the original complement strand includes a minimum of 3 members in each set.

65. (Previously Presented) The method of claim 55, wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject, and wherein the method further comprises:

comparing the error-corrected consensus sequence to a respective reference sequence; and

detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

66. (Previously Presented) The method of claim 65, further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

REMARKS

Claims 37-66 were pending when the present Restriction Requirement was mailed on August 28, 2020. In this response, claims 50, 54-57, 59, and 61 are amended solely for clarity purposes, and are made without prejudice to pursuing these claims in unamended or other forms in a continuation or other application. These amendments do not add any new matter. Upon entry, claims 37-66 remain pending.

In response to the restriction requirement mailed August 28, 2020, Applicant elects the following species, without traverse:

1. Species of variations present in the consensus sequence are indicative of: (claim 51)

A) a cancer risk

2. Species of location of molecule-specific feature: (claim 53)

H) molecule-specific feature is at an end of the dsDNA molecule

3. Species of adapter molecule: (claim 57)

L) comprises a double-stranded SMI sequence

4. Species of adapter molecule: (claim 59)

M) comprises a Y-shape

5. Species of set of duplicates of original forward and complement strands: (claims 63 and 64)

0) includes a minimum of 1 member in each set

6. Species of source of duplex DNA molecules in a subject: (claim 65)

Q) derived from a tumor

Independent claims 37, 46 and 55 are generic to all elected species, and at least claims 37-63 and 65-66 encompass the elected species.

Dated: September 29, 2020

Respectfully submitted,

Customer No. 94991
UW/Perkins Coie LLP
Patent - LA
P.O. Box 1247
Seattle, WA 98111-1247
Phone: (310) 788-3349
Fax: (206) 332-7198

PERKINS COIE LLP

By: /Lara J. Dueppen/
Lara J. Dueppen
Registration No. 65,002

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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]
	C1	Borodina, T. et al. "Chapter five - A Strand-Specific Library Preparation Protocol for RNA Sequencing" Methods in Enzymology, 2011, Vol. 500, pp. 79-98.	
	C2	USPTO, Application No. 16/411,066, Non-Final Office Action, Examiner David C. THOMAS, mail date: 12/23/2019. 12 pages.	
	C3	USPTO, Application No. 16/411,066, Notice of Allowance, Examiner David C. THOMAS, mail date: 3/6/2020. 6 pages.	
	C4	USPTO, Application No. 16/411,068, Notice of Allowance, Examiner David C. THOMAS, mail date: 2/28/2020. 6 pages.	
	C5	USPTO, Application No. 16/411,069, Notice of Allowance, Examiner David C. THOMAS, mail date: 3/6/2020. 6 pages.	
	C6	USPTO, Application No. 16/514,931, Non-Final Office Action, Examiner David C. THOMAS, mail date: 1/29/2020. 12 pages.	
	C7	USPTO, Application No. 16/514,931, Notice of Allowance, Examiner David C. THOMAS, mail date: 6/1/2020. 5 pages.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C8	USPTO, Application No. 16/503,382, Non-Final Office Action, Examiner Stephanie K. MUMMERT, mail date: 1/24/2020. 12 pages.	
	C9	USPTO, Application No. 16/503,382, Notice of Allowance with notice of references cited, Examiner Stephanie K. MUMMERT, mail date: 7/15/2020. 12 pages.	
	C10	USPTO, Application No. 16/503,382, Corrected Notice of Allowance with notice of references cited, Examiner Stephanie K. MUMMERT, mail date: 7/29/2020. 4 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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2020-09-29
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Electronic Acknowledgement Receipt

EFS ID:	40711032
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	29-SEP-2020
Filing Date:	30-AUG-2018
Time Stamp:	20:38:01
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		2020-09-29-Response-to-Restriction-072227-8043US07.pdf	107009 bbf3228de3f24cb7e408923df370ee2e4b7cd92c	yes	11

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Response to Election / Restriction Filed	1	1
	Claims	2	9
	Applicant Arguments/Remarks Made in an Amendment	10	11

Warnings:					
Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	2020-09-29-Supp-IDS-072227-8043US07.pdf	224542 7eaf72030e827671f5f6634c6aa8456411703367	no	3
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
3	Non Patent Literature	C01-Borodina-MiE-2011.PDF	1376598 1072e147a74054cf490b9751a54388da4955b3db	no	20
Warnings:					
Information:					
4	Non Patent Literature	C02-US16411066-2019-12-23-NFOA-8043US13.PDF	2216012 0a02f09cf784e4f200876b06f7f822f9804993a0	no	12
Warnings:					
Information:					
5	Non Patent Literature	C03-US16411066-2020-03-06-NOA-8043US13.PDF	1090279 e2a54305220f52461931e8b2f428b91763d85f6f	no	6
Warnings:					
Information:					
6	Non Patent Literature	C04-US16411068-2020-02-28-NOA-8043US14.PDF	923943 5a8ba5dcd959c6bdbbf0b9cb5ba5f06847145d04	no	6
Warnings:					
Information:					

7	Non Patent Literature	C05-US16411069-2020-03-06-NOA-8043US15.PDF	924787 654caa6d1e323eadefa66936369845976d096aee	no	6
Warnings:					
Information:					
8	Non Patent Literature	C06-US16514931-2020-01-29-NFOA-8043US16.PDF	538872 065d5e278a00278da93ecfdc41deec3201317b14	no	12
Warnings:					
Information:					
9	Non Patent Literature	C07-US16514931-2020-06-01-NOA-8043US16.PDF	918690 81f6955628a69d4a43621601801a8db3f257565a	no	5
Warnings:					
Information:					
10	Non Patent Literature	C08-US16503382-2020-01-24-NFOA-8043US17.PDF	2703644 659092a8697ae339814c63f386a8d220d056c081	no	12
Warnings:					
Information:					
11	Non Patent Literature	C09-US16503382-2020-07-15-NOA-8043US17.PDF	2726557 d4d7a46ca7e31df1fc2bcce254b86d90dde6dfff	no	12
Warnings:					
Information:					
12	Non Patent Literature	C10-US16503382-2020-07-29-cNOA-8043US17.PDF	849207 17814b0465a2f355866941c52231b5224fda4185	no	4
Warnings:					
Information:					
			Total Files Size (in bytes):	14600140	

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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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
	A1	2014/0155274	A1	2014-06-05	Xie et al.	

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]
	C1	EPO, Extended European Search Report for Application No. EP20166917.3, dated September 30, 2020. 6 pages.	
	C2	USPTO, Application No. 16/908,611, Non-Final Office Action, Examiner David C. THOMAS, mail date: 10/9/2020. 28 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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2020-10-24
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Electronic Acknowledgement Receipt

EFS ID:	40952117
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	26-OCT-2020
Filing Date:	30-AUG-2018
Time Stamp:	22:02:14
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	Information Disclosure Statement (IDS) Form (SB08)	2020-10-26-Supp-IDS-072227-8043US07.pdf	70120 4980fb2f76204bbd7040b836c498be0da62db988	no	2

Warnings:

Information:					
This is not an USPTO supplied IDS fillable form					
2	Non Patent Literature	C01-2020-09-30-EESR-072227-8043EP00.PDF	4945649 e56190dac8a1b3dc46db8f0d8584bc1624993c51	no	6
Warnings:					
Information:					
3	Non Patent Literature	C02-US16908611-2020-10-09-NFOA-072227-8043US19.PDF	2419463 afadf6719df178a2f08689a5be765bb67c7f6c4c	no	28
Warnings:					
Information:					
			Total Files Size (in bytes):	7435232	

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.

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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

94991 7590 11/18/2020

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
11/18/2020	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):

patentprocurement@perkinscoie.com

Office Action Summary	Application No. 16/118,306	Applicant(s) Salk et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status No

-- 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 29 September 2020.

☐ 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) 37-66 is/are pending in the application.

5a) Of the above claim(s) 64 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 37-63 and 65-66 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 _____ 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 10/26/2020, 9/29/2020, 12/16/2019, 6/21/2019.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions.

Election/Restrictions

2. Applicant's election without traverse of various species in the reply filed on September 29, 2020 is acknowledged. Applicant has elected the species A) a cancer risk, from the group of species of variations present in the consensus sequence are indicative of, drawn to claim 51, the species H) molecule-specific feature is at an end of the dsDNA molecule, from the group of species of location of molecule-specific feature, drawn to claim 53, the species L) comprises a double-stranded SMI sequence, from the group of species of adapter molecule, drawn to claim 57, the species M) comprises a Y-shape, from the group of species adapter molecule, drawn to claim 59, the species O) includes a minimum of 1 member in each set, from the group of species of duplicates of original forward and complement strands, drawn to claims 63 and 64, and the species Q) derived from a tumor, from the group of species of source of duplex DNA molecules in a subject, drawn to claim 65. Claims 37, 46 and 55 are generic to the elected species, which read on claims 37-63, 65 and 66, which will be examined on the merits. Claim 64, drawn to non-elected species, is currently withdrawn. Claims 1-36 were previously canceled. It is also noted that claims 50, 54-57, 59 and 61 are currently amended.

Double Patenting

Nonstatutory Obviousness-type Double Patenting Rejections

3. 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. See *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); and *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) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

With respect to the non-statutory double patenting rejection(s) made in this Office action, note as follows. The use of the terminology "defined in 35 U.S.C. §154 to §156 and §173" in a terminal disclaimer can result in the terminal disclaimer being found improper. To address this, note that a proper terminal disclaimer need only disclaim the patent's remaining "full statutory term" as defined in 35 U.S.C., without specifying 35 U.S.C. 154 and 173. This is so, because the "full statutory term" inherently is a statutorily defined item.

The USPTO internet Web site contains terminal disclaimer forms which may be used. Please visit http://www.uspto.gov/forms/. The filing date of the application will determine what form 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 http://www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

4. Claims 37-63, 65 and 66 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-23 of U.S. Patent No. 10,287,631 (cited on IDS of 12/16/2019). Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims of U.S. Patent No. 10,287,631 represents a narrower species of the current claims and this

species anticipates the current, more generic claims. Both sets of claims teach methods of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising the steps of amplifying each original strand of the double-stranded target nucleic acid molecule, sequencing the amplified target nucleic acid products generated from each strand, confirming the presence of at least one sequence read from each of the original strands, comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of both strands is complementary. The claims of U.S. Patent No. 10,287,631 also teach that the methods comprise ligating each of the double-stranded target nucleic acid molecules to at least one adapter molecule that comprises a degenerate or semi-degenerate single molecule identifier (SMI) sequence and a strand-distinguishing nucleotide sequence, grouping the first and second strand sequence reads into a family of first and second strand sequence reads based on the SMI sequence alone or in combination with the target nucleic acid fragment ends, separating the first and second strand sequence reads into a set of first strand sequence reads and a set of second strand sequence reads based on the region of non-complementarity between the first and second strand of the adapter-target nucleic acid complex, and identifying nucleotide positions where the compared first and second strand sequence reads are non-complementary and where they are complementary.

5. Claims 37-63, 65 and 66 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/118,290. Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims of copending U.S. Application No. 16/118,290 represents a narrower species of the current claims and this species anticipates the current, more generic claims. Both sets of claims teach methods of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising the steps of amplifying each original strand of the double-stranded target nucleic acid molecule, sequencing the amplified target nucleic acid products generated from each strand, confirming the presence of at least one sequence read from each of the original strands (see claim 41 of the '290 application), comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of both strands is complementary (see claim 38 of the '290 application). The claims of copending U.S. Application No. 16/118,290 also teach that the methods comprise attaching adapters to both ends of double-stranded nucleic acid fragments in a sample to provide adapter-nucleic acid products, wherein the adapters comprise a double-stranded hybridized region, single-stranded 5' and 3' arms, and a physical unique molecular identifier (UMI) on at least one of the single-stranded arms that can be used to identify an individual molecule of a double-stranded nucleic acid fragment in the sample, sequencing the plurality of amplified polynucleotides to obtain a plurality of sequence reads each

associated with a physical UMI, identifying a plurality of physical UMIs associated with the plurality of reads, and determining high accuracy sequence reads of the double-stranded nucleic acid fragments in the sample using the plurality of sequences obtained and the plurality of physical UMIs identified.

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

6. Claims 37-63, 65 and 66 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/503,398. Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims of copending U.S. Application No. 16/503,398 represents a narrower species of the current claims and this species anticipates the current, more generic claims. Both sets of claims teach methods of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising the steps of amplifying each original strand of the double-stranded target nucleic acid molecule (see claim 40 of the '398 application), sequencing the amplified target nucleic acid products generated from each strand, confirming the presence of at least one sequence read from each of the original strands (see claim 55 of the '398 application), comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of both strands is complementary. The claims of copending U.S.

Application No. 16/503,398 also teach that the methods comprise preparing a sequencing library from a sample comprising a plurality of double-stranded DNA fragments by ligating adapter molecules to the plurality of double-stranded DNA fragments to generate a plurality of double-stranded adapter-DNA molecules).

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

Analysis of Subject Matter Eligibility under 35 U.S.C. 101

7. It is noted that independent claims 37, 46 and 55 generally recite the active steps of amplifying each original strand of the double-stranded target nucleic acid molecule, and sequencing the amplified target nucleic acid products generated from each strand to generate sequencing reads from each of the original strands (claim 37), sequencing first and second strands of a double-stranded DNA molecule to generate first and second strand sequencing reads (claim 46), and ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules, for each of one or more tagged molecules generating a set of duplicates of an original forward strand and an original complement strand, and creating a single-strand consensus sequence (SSCS) from both sets of duplicates (claim 55). Claim 37 also recites the steps of confirming the presence of at least one sequence read of an amplified target nucleic product generated from each of the original strands, comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of

both strands is complementary. Claim 46 also recites the steps of grouping the first and second sequence reads by a common feature, comparing the first and second sequencing reads to each other, and generating a consensus sequence of nucleotide bases wherein each nucleotide base is identified at a given position in the consensus sequence when a specific nucleotide is complementary between the first strand sequencing read and the second strand sequencing read. Claim 55 also recites the steps of comparing the SSCS of the original forward and complement strands to each other, and generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward and complement strands are complementary.

Since the claims recite limitations that potentially include judicial exceptions in the form of abstract ideas relating to comparison of nucleic acid sequences and sequence data analysis to generate high accuracy sequence reads of nucleic acid molecules, claims 37-63, 65 and 66 were reviewed with regard to subject matter eligibility under 35 U.S.C. 101 based on the 2019 Revised Patent Subject Matter Eligibility Guidance (published January, 2019 and updated, October, 2019).

Based on analysis of the claims using the revised guidelines, the claims are drawn to a process and are therefore directed to a statutory category in step 1. In step 2, a two-prong inquiry is performed to determine if a claim is directed to a judicial exception wherein in the first prong, in the case of potential abstract ideas, limitations are evaluated to determine if the claims fall within the enumerated groupings of abstract ideas by reciting a mathematical concept, a mental process or certain methods of organizing human activity. In the current claims, the steps of grouping and comparing

sequence reads, and generating consensus sequences based on the comparison of the sequence reads, could, in some situations, be performed by a computer if the number of target molecules and sequence reads is substantially large. However, without further evidence, since there is no indication that the double-stranded target nucleic acid molecules or double-stranded DNA molecules is a large pool of molecules (in fact in claims 37 and 46, single molecules are sequenced), or how many sequence reads are generated, the processing of a relatively small number of double-stranded nucleic acid or DNA molecules to generate a high accuracy sequence read or an error-corrected sequence read is a mental process that could conceivably be performed in the human mind, and therefore these limitations fall within the enumerated groupings of abstract ideas and are considered judicial exceptions. Therefore, in the second prong of step 2, the claims are evaluated to determine if there are any additional elements recited in the claims beyond the judicial exceptions, and evaluating those elements to determine whether they integrate the exceptions into a practical application of the exception. It is determined that the claims integrate the judicial exceptions into a practical application since the claimed invention pertains to an improvement in the field of nucleic acid sequence analysis and generating high accuracy sequences, as the method allows for the generating accurate and error-corrected consensus sequences which reduces or eliminates artifactual mutations arising from DNA damage, PCR errors or sequencing errors, and allow rare variants in heterogeneous populations to be detected (see paragraphs 8-11 and 105 of the specification). Therefore, the invention as claimed qualifies as eligible subject matter.

Claim Objections

8. Claims 55 and 59 are objected to because of the following informalities: The last word recited in claim 55 is “complimentary”. Even though this is an accepted spelling of this term, “complementary” is also acceptable, and since other uses of the term in the claims are recited as “complementary”, all uses should read “complementary” for consistency. Claim 59 recites the limitation “wherein the adapter molecules comprises a Y-shape, a U-shape, or a combination thereof”. To be grammatically correct, the limitation should read “wherein the adapter molecules comprise a Y-shape, a U-shape, or a combination thereof”. Appropriate correction is required.

Claim Rejections - 35 USC § 112

9. The following is a quotation of 35 U.S.C. 112 (pre-AIA), second paragraph:
The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

10. Claim 58 is rejected under 35 U.S.C. 112 (pre-AIA), second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention. Regarding claim 58, the phrase "and preferably" renders the claim indefinite because it is unclear whether the limitation(s) following the phrase are part of the claimed invention. See MPEP § 2173.05(d).

Claim Rejections - 35 USC § 102

11. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any

correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

12. The following is a quotation of the appropriate paragraphs of pre-AIA 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

13. Claim(s) 37-39, 41, 42, 45-48, 52-58 and 60-63 is/are rejected under pre-AIA 35 U.S.C. 102(b) as being anticipated by Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019).

With regard to claim 37, Travers teaches a method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule (methods are provided for sequencing nucleic acid template constructs comprising double-stranded portions to provide redundant sequencing of both the sense and antisense strands, see Abstract and paragraph 9) comprising:

amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products (template constructs may be prepared from PCR products to provide a population of identical molecules from which to generate sequence reads, paragraph 54, lines 12-16, paragraph 66, lines 1-9 and paragraph 88, lines 1-4; template constructs may also be amplified by a rolling circle replication

process to generate repeated copies of the template to provide for sequencing redundancy, paragraph 122, lines 1-21);

sequencing the amplified target nucleic acid products generated from each original strand (the method comprises identifying the nucleotide sequences of both strands of each double-stranded nucleic acid molecule starting from each end, paragraph 13, lines 1-10, paragraph 15, lines 5-17, paragraph 54, lines 12-16 and paragraph 88, lines 4-7);

confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands (the progress of the sequencing of both strands is followed, wherein a primer is extended on one strand and continues around a connecting hairpin to sequence to opposite strand, paragraph 54, lines 16-33 and Figure 3A; see also paragraphs 139-141 and Figure 11 showing a sequencing trace of 700 bases representing both strands of a duplex segment on a circular hairpin template, including 154 baes of consensus information);

comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand (multiple reads of the nucleic acid sequence of the sense and antisense strands are compared, paragraph 15, lines 17-20 and paragraph 57, lines 14-17); and

generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary (following comparison of

multiple sequence reads of each strand, a consensus sequence is established wherein each base call in a sequence is based on a consensus base call for that position based upon multiple reads at that position and/or reads at that position and its complement in the antisense strand, paragraph 15, lines 17-20, paragraph 57, lines 11-17 and paragraph 58, lines 1-14).

With regard to claim 38, Travers teaches a method wherein prior to amplifying each strand, the method further comprises ligating the double-stranded target nucleic acid molecule to an adapter molecule resulting in an asymmetric adapter-target nucleic acid complex such that each original strand generates a distinct yet related set of amplified target nucleic acid products during the amplification step (a single template construct may comprise asymmetric linking oligonucleotides for ligating adapter/linking hairpin sequences, and therefore the adapter/linking hairpin sequences ligated to each end are also asymmetric, paragraph 92, lines 13-24).

With regard to claims 39, 54, and 56-58, Travers teaches a method wherein prior to sequencing either the first strand or the second strand, the method further comprises ligating the dsDNA molecule to at least one adapter molecule to form an adapter-dsDNA complex, and wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule that provides a common feature comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence, wherein the SMI adapter molecules comprise a single-stranded SMI sequence or a double-stranded SMI sequence (template constructs may comprise linking oligonucleotides that are ligated to each end of a nucleic acid fragment that include identifiable sequence characteristics such as barcodes, generally comprising random or degenerate

nucleotides, to facilitate its identification with respect to the contiguous template sequence with which it is associated and to other template sequences present in the same sample, and thus the origin of each sequence is determined following sequencing of each template and its associated identifiable sequence characteristic, paragraph 16, lines 1-18, and also paragraphs 100-102 and Figure 7 showing ligation of the hairpin adapters to the linking oligonucleotides, and thus the sequences comprising the linking oligonucleotides become double-stranded).

With regard to claim 41, Travers teaches a method wherein the confirming step includes (i) grouping the sequence reads of the amplified target nucleic acid products into families of paired target nucleic acid strand sequences using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the double-stranded target nucleic acid molecule as an identifier, and (ii) separating each individual family of paired target nucleic acid strand sequences into a set of sequence reads from one original strand of the double-stranded target nucleic acid molecule and a set of sequence reads from the other original strand of the double stranded target nucleic acid molecule (template constructs may comprise linking oligonucleotides that include identifiable sequence characteristics such as barcodes to facilitate its identification with respect to the contiguous template sequence with which it is associated and to other template sequences present in the same sample, and thus the origin of each sequence is determined following sequencing of each template and its associated identifiable sequence characteristic and analyzed separately from sequence reads of other, distinct template nucleic acids in the sample, paragraph 16, lines 1-18, and also paragraphs 100-102).

With regard to claim 42, Travers teaches a method wherein the confirming step includes confirming the presence of at least about 3 sequence reads of an amplified target nucleic acid product generated from each of the original strands, and wherein the comparing step includes comparing a first consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from one original strand with a second consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from the other original strand (sequencing is typically performed at greater than 20x coverage to provide sequence accuracy of over 99.999%, paragraph 147, lines 1-8).

With regard to claims 45 and 47, Travers teaches a method wherein generating a consensus sequence of the double-stranded target nucleic acid molecule further comprises identifying one or more nucleotide positions that disagree between the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand, and discounting nucleotide positions in the consensus sequence that are in disagreement to provide an error-corrected sequence read of the double-stranded target nucleic acid molecule (algorithms are used that classify the sequence read data into a consensus basecall for each position, often with an associated error, wherein multiple steps are performed to reduce he possibility of errors between reads form a single molecule or between reads from different molecules, paragraph 60, 1-30).

With regard to claim 46, Travers teaches a method for generating a high accuracy DNA sequencing read (methods are provided for sequencing nucleic acid

template constructs comprising double-stranded DNA portions to provide redundant sequencing of both the sense and antisense strands, see Abstract, paragraph 9 and paragraph 89, lines 1-4), comprising:

sequencing a first strand of a double-stranded DNA (dsDNA) molecule to generate a first strand sequencing read; sequencing a second strand of the same dsDNA molecule to generate a second strand sequencing read (the method comprises identifying the nucleotide sequences of both strands of each double-stranded nucleic acid molecule starting from each end, paragraph 13, lines 1-10, paragraph 15, lines 5-17, paragraph 54, lines 12-16 and paragraph 88, lines 4-7);

grouping the first and second sequence reads by a common feature (template constructs may comprise linking oligonucleotides that include identifiable sequence characteristics such as barcodes to facilitate its identification with respect to the contiguous template sequence with which it is associated and to other template sequences present in the same sample, and thus the origin of each sequence is determined following sequencing of each template and its associated identifiable sequence characteristic, paragraph 16, lines 1-18, and also paragraphs 100-102);

comparing the first strand sequencing read to the second strand sequencing (multiple reads of the nucleic acid sequence of the sense and antisense strands are compared, paragraph 15, lines 17-20 and paragraph 57, lines 14-17); and

generating a consensus sequence comprising a sequence of nucleotide bases, wherein each nucleotide base is identified at a given position in the consensus sequence when a specific nucleotide is complementary between the first strand sequencing read and the second strand sequencing read (following comparison of

multiple sequence reads of each strand, a consensus sequence is established wherein each base call in a sequence is based on a consensus base call for that position based upon multiple reads at that position and/or reads at that position and its complement in the antisense strand, paragraph 15, lines 17-20, paragraph 57, lines 11-17 and paragraph 58, lines 1-14).

With regard to claim 48, Travers teaches a method further comprising comparing the consensus sequence to a respective reference sequence, and identifying one or more nucleotide sequence variations present in the consensus sequence not present in the reference sequence (following collection of sequence data for a particular set of constructs, the sequences are compared to a reference sequence, paragraph 146, lines 29-32).

With regard to claims 52 and 53, Travers teaches a method wherein the common feature comprises one or more DNA molecule features that distinguish the dsDNA molecule from other dsDNA molecules in a sample, wherein the one or more DNA molecule features includes a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the dsDNA molecule, or a combination thereof (connecting or linking oligonucleotides are selected and ligated to the ends of double-stranded fragments and comprise identifiable sequence characteristics to distinguish a particular construct from constructs generated for other templates, paragraph 100, lines 1-16 and Figure 7).

With regard to claim 55, Travers teaches a method of generating an error-corrected consensus sequence (methods are provided for sequencing nucleic acid template constructs comprising double-stranded portions to provide redundant

sequencing of both the sense and antisense strands, including the use of algorithms to minimize errors in base calling, see Abstract and paragraphs 9 and 60) comprising:

ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules (exogenous hairpin adapter sequences are ligated to the ends of double-stranded fragments, paragraph 18, lines 1-12, paragraph 51, lines 12-20 and Figure 7);

for each of one or more tagged DNA molecules

generating a set of duplicates of an original forward strand of the tagged DNA molecule and a set of duplicates of an original complement strand of the tagged DNA molecule (template constructs may be prepared from PCR products to provide a population of identical molecules from which to generate sequence reads, paragraph 54, lines 12-16, paragraph 66, lines 1-9 and paragraph 88, lines 1-4; template constructs may also be amplified by a rolling circle replication process to generate repeated copies of the template to provide for sequencing redundancy, paragraph 122, lines 1-21);

creating a single strand consensus sequence (SSCS) from the set of duplicates of the original forward strand and a single strand consensus sequence (SSCS) from the set of duplicates of the original complement strand (consensus sequences are first generated for the sense strand and the antisense strand of a duplex, paragraph 59, lines 1-5 and paragraph 61, lines 1-11);

comparing the SSCS of the original forward strand to the SSCS of the original complement strand (the two consensus sequences are then compared using any of various methods and algorithms, paragraph 61, lines 11-18); and

generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward strand and the SSCS of the original complement strand are complimentary (the consensus sequences are then combined, such as with a consensus determination algorithm, paragraph 61, lines 11-18).

With regard to claim 60, Travers teaches a method wherein the tagged DNA molecule comprises an adapter molecule at each end (exogenous hairpin adapter sequences are ligated to both ends of double-stranded fragments after first attaching linking oligonucleotides to the fragments, paragraph 51, lines 12-20 and Figure 7).

With regard to claim 61, Travers teaches a method wherein the comparing step includes pairing the SSCS of the original forward strand to the SSCS of the original complement strand using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of a particular duplex DNA molecule as an identifier (after consensus sequences are established for each strand of a fragment, the two consensus sequences are then compared using any of various methods and algorithms, wherein sequences of the linking oligonucleotides at each end allow the sequences to be paired, paragraph 61, lines 1-18 and paragraph 88, lines 1-7).

With regard to claim 62, Travers teaches a method further comprising sequencing one or more of the duplicates of the original forward strand and one or more of the duplicates of the original complement strand of the tagged DNA molecules using a next-generation sequencing procedure (sequencing may include the use of systems provided by Illumina, Inc. or 454 Life Sciences, paragraph 133, lines 1-14).

With regard to claim 63, Travers teaches a method wherein the set of duplicates of the original forward strand and the set of duplicates of the original complement strand includes a minimum of 1 member in each set (multiple copies of the fragment are generated prior to sequencing to provide multiple copies of both the sense and antisense strands, paragraph 122, lines 1-21).

Claim Rejections - 35 USC § 103

14. In the event the determination of the status of the application as subject to AIA 35 U.S.C. 102 and 103 (or as subject to pre-AIA 35 U.S.C. 102 and 103) is incorrect, any correction of the statutory basis for the rejection will not be considered a new ground of rejection if the prior art relied upon, and the rationale supporting the rejection, would be the same under either status.

15. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

16. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to

consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

17. Claims 40 and 59 is/are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019) in view of Xie et al. U.S. Patent Pub. No. 2014/0155274, cited on IDS of 10/26/2020).

Travers teaches the limitations of claims 37-39, 41, 42, 45-48, 52-58 and 60-63, as discussed above. However, Travers does not teach a method wherein the adapter molecule further comprises a nucleotide sequence that uniquely tags each original strand of the double-stranded target nucleic acid molecule such that each original strand is distinctly identifiable relative to its original complementary strand, wherein the adapter molecules comprise a Y-shape, a U-shape, or a combination thereof.

Xie teaches methods for identifying target nucleic acid molecules in a sample, such as the plurality of target molecules from a single cell, using unique barcode sequences for tagging each nucleic acid molecule in the sample (see paragraph 7).

Specifically with regard to claims 40 and 59, Xie teaches methods of digital counting by random labeling of all target nucleic acid molecules in a sample with unique barcode sequences, wherein barcoded phosphorylated Illumina Y-shaped adapters are ligated to both ends of cDNA fragments in a sample, wherein the sample is then amplified by PCR and prepared for sequencing using the standard Illumina library protocol, wherein for each amplicon, both barcoded sequences and both strands of the

cDNA sequence are read using paired end deep sequencing (see paragraph 138 and Figure 10A and B).

It would have been prima facie obvious to one having ordinary skill in the art at the time the invention was made to combine the methods of Travers and Xie since both references teach methods for nucleic acid sequence analysis comprising labeling double-stranded molecules of a sample with double-stranded adapters comprising unique identifiable sequence characteristics or barcode sequences such that each molecule is labeled with a unique barcode sequence that is used in paired end sequencing such that the barcode sequence links each strand with its original complementary strand. While Travers teaches that target nucleic acid molecules are ligated to linker oligonucleotides and hairpin adapters to link the two strands of a double-stranded fragment prior to sequencing (see Travers, paragraph 16 and Figure 7), Xie teaches that Y-shaped barcode adapters are ligated onto both ends of the target molecules prior to amplification and cloning the adapter-target molecules to form a standard Illumina sequencing library (see Xie, paragraph 138 and Figure 10). Thus, an ordinary practitioner would have been motivated to combine the methods of Travers and Xie since Xie provides another means for labeling target nucleic acid molecules with unique barcode molecules to perform paired end sequencing. Xie further teaches that "attaching a barcode to each end of the nucleic acid in a sample, reduces the number of barcodes that must be designed and synthesized while allowing conventional paired-end library protocols and providing long-range sequence information that improves mapping accuracy".

18. Claims 43, 44, 49-51, 65 and 66 is/are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019) in view of Bielas et al. (U.S. Patent Pub. No. 2015/0024950, cited on IDS of 6/21/2019).

Travers teaches the limitations of claims 37-39, 41, 42, 45-48, 52-58 and 60-63, as discussed above.

In addition, with regard to claim 51, Travers teaches a method wherein the dsDNA molecule is one of a population of dsDNA molecules, and wherein the method further comprises generating a consensus sequence for a plurality of the dsDNA molecules (template constructs may be prepared from PCR products to provide a population of identical molecules from which to generate sequence reads, paragraph 54, lines 12-16, paragraph 66, lines 1-9 and paragraph 88, lines 1-4; template constructs may also be amplified by a rolling circle replication process to generate repeated copies of the template to provide for sequencing redundancy, paragraph 122, lines 1-21; consensus sequences are generated for each of the sense and antisense strands, and the consensus sequences are then combined to generate a consensus sequence for the double-stranded template (see paragraph 61), and wherein for each of one or more of the generated consensus sequences the method further comprises:

comparing the consensus sequence to a reference sequence (following collection of sequence data for a particular set of constructs, the sequences are compared to a reference sequence, paragraph 146, lines 29-32).

However, with regard to claims 43, 44 and 49, Travers does not teach a method further comprising identifying a mutation or variation occurring at a particular position in

the consensus sequence as a true mutation, and further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation. With regard to claim 50, Travers also does not teach a method further comprising identifying a potential artifact at a particular position when a base call in the first strand sequence read does not agree with the base call in the second strand sequence read. With regard to claim 51, Travers also does not teach a method that comprises identifying one or more variations present in the consensus sequence not present in the reference sequence, wherein the one or more variations present in the consensus sequence are indicative of a cancer risk, a cancer metabolic state, a mutator phenotype, a carcinogen exposure, a chronic inflammation exposure, an age, a neurodegenerative disease, or a combination thereof in an organism from which the dsDNA molecule is derived. With regard to claims 65 and 66, Travers also does not teach a method wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject, and wherein the method further comprises comparing the error-corrected consensus sequence to a respective reference sequence and detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence, or a method further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

Bielas teaches methods for detecting mutations in double-stranded nucleic acid molecules by massively parallel sequencing methods, wherein the target molecules are

ligated into a cypher vector library that comprises the random cyphers and adapter sequences (see Abstract and paragraph 8, lines 1-9, paragraph 12, lines 1-4, paragraph 13, lines 7-10 and Figure 2). Bielas further teaches that the sequences of both strands of target sequence associated with first and second random cyphers are aligned to establish consensus sequences (paragraph 10, lines 10-15, paragraph 13, lines 11-19, paragraph 79, lines 1-11 and Figure 3B)

Specifically with regard to claims 43, 44 and 49, Bielas teaches a method further comprising identifying a mutation or variation occurring at a particular position in the consensus sequence as a true mutation (the sequencing results obtained using the dual cypher library for a TP53 Exon 4 DNA sequence were compared to sequences of a wild-type TP53 Exon DNA sample to distinguish a true mutation that arose during cell replication from errors generated during processing steps after barcoding, such as during PCR or sequencing, paragraph 14, lines 1-7, paragraph 80, lines 1-16 and Figure 4A and B), and further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation (the dual cypher sequencing method may be applied to normal and tumor genomic libraries, including detection of rare mutants against a large background signal, wherein genomic DNA is prepared from patient-matched normal and tumor tissue samples, or may be used for detecting circulating mutant DNA in blood, paragraph 64, lines 1-7 and paragraph 66, lines 1-8).

With regard to claim 50, Bielas teaches a method further comprising identifying a potential artifact at a particular position when a base call in the first strand sequence read does not agree with the base call in the second strand sequence read (errors

occurring as a result of PCR or sequencing are detected as arising in one strand but not the complementary strand, paragraph 16, lines 13-22).

With regard to claim 51, Bielas teaches

comparing the consensus sequence to a reference sequence and identifying one or more variations present in the consensus sequence not present in the reference sequence (the sequencing results obtained using the dual cypher library for a TP53 Exon 4 DNA sequence were compared to sequences of a wild-type TP53 Exon DNA sample to distinguish a true mutation that arose during cell replication from errors generated during processing steps after barcoding, such as during PCR or sequencing, paragraph 14, lines 1-7, paragraph 80, lines 1-16 and Figure 4A and B); and

wherein the one or more variations present in the consensus sequence are indicative of a cancer risk, a cancer metabolic state, a mutator phenotype, a carcinogen exposure, a chronic inflammation exposure, an age, a neurodegenerative disease, or a combination thereof in an organism from which the dsDNA molecule is derived (the methods may be used for quantifying DNA damage that my indicative of response to therapy, or for monitoring disease progression, or for monitoring DNA damage resulting from chemotherapy, or in detecting and quantitation of specific methylation of DNA sequences, which may require repair of treated DNAs as controls, paragraph 64, lines 1-15).

With regard to claims 65 and 66, Bielas teaches a method wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject (the dual cypher sequencing method may be used for detecting circulating

mutant DNA in blood, paragraph 64, lines 1-7), and wherein the method further comprises:

comparing the error-corrected consensus sequence to a respective reference sequence (the sequencing results obtained using the dual cypher library for a TP53 Exon 4 DNA sequence were compared to sequences of a wild-type TP53 Exon DNA sample to distinguish a true mutation that arose during cell replication from errors generated during processing steps after barcoding, such as during PCR or sequencing, paragraph 14, lines 1-7, paragraph 80, lines 1-16 and Figure 4A and B); and

detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence, and further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence (the dual cypher sequencing method may be applied to monitoring or detecting disease- and drug response-associated mutations, paragraph 64, lines 1-7).

It would have been prima facie obvious to one having ordinary skill in the art at the time the invention was made to combine the methods of Travers and Bielas since both references teach methods for nucleic acid sequence analysis comprising labeling double-stranded molecules of a sample with double-stranded adapters comprising unique identifiable sequence characteristics or barcode sequences such that each molecule is labeled with a unique barcode sequence that is used in paired end sequencing such that the barcode sequence links each strand with its original complementary strand. While Travers teaches that target nucleic acid molecules are

ligated to linker oligonucleotides and hairpin adapters to link the two strands of a double-stranded fragment prior to sequencing (see Travers, paragraph 16 and Figure 7), Bielas teaches that target nucleic acid molecules are ligated into a cypher vector library comprising previously catalogued double-stranded cyphers prior to amplification and sequencing (see Bielas, paragraph 13). Thus, an ordinary practitioner would have been motivated to combine the methods of Travers and Bielas since Bielas provides another means for labeling target nucleic acid molecules with unique barcode molecules to perform paired end sequencing, also teaches methods for comparing consensus sequences obtained for double-stranded template sequences to each other to identify amplification or sequencing artifacts, and comparing the consensus sequences to a reference sequence to identify true mutations. In addition, Bielas teaches additional types of samples that may be tested using the methods of Travers, including samples derived from tumors or circulating nucleic acids, or applying the methods to monitor disease progression or response to therapy (see Bielas, paragraph 64, lines 1-15 and paragraph 66, lines 1-8). Furthermore, the methods of Travers that uses contiguous or circular template sequences offer the advantages for obtaining enhanced and iterative replicate sequence data for both sense and antisense strands (see paragraphs 55, 63 and 64).

Conclusion

19. Claims 37-63, 65 and 66 are rejected are rejected over the prior art. No claims are free of the prior art. In addition, claims 37-63, 65 and 66 are rejected on the grounds of nonstatutory obviousness-type double patenting, as discussed above.

Correspondence

20. Any inquiry concerning this communication or earlier communications from the examiner should be directed to David C. Thomas whose telephone number is 571-272-3320 and whose fax number is 571-273-3320. The examiner can normally be reached on 5 days, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on 571-272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/David C Thomas/
Primary Examiner, Art Unit 1637

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

CPC - Searched*		
Symbol	**Date**	**Examiner**
C12Q 1/6806	11/10/2020	DCT
C12Q 1/6876	11/10/2020	DCT
C12Q 1/6869	11/10/2020	DCT

CPC Combination Sets - Searched*		
Symbol	**Date**	**Examiner**

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**
EAST	11/09/2020	DCT
EAST	11/10/2020	DCT
EAST	11/12/2020	DCT
Google-search terms: sequencing double-stranded dna, both strands , barcodes, tags, y-shaped adapters, consensus sequences, sequence reads, single molecule identifier sequences, smi adapters, single strand consensus sequences	11/09/2020	DCT
PALM, inventor name search	11/09/2020	DCT

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

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

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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]
/D.C.T/	C1	Borodina, T. et al. "Chapter five - A Strand-Specific Library Preparation Protocol for RNA Sequencing" Methods in Enzymology, 2011, Vol. 500, pp. 79-98.	
/D.C.T/	C2	USPTO, Application No. 16/411,066, Non-Final Office Action, Examiner David C. THOMAS, mail date: 12/23/2019. 12 pages.	
/D.C.T/	C3	USPTO, Application No. 16/411,066, Notice of Allowance, Examiner David C. THOMAS, mail date: 3/6/2020. 6 pages.	
/D.C.T/	C4	USPTO, Application No. 16/411,068, Notice of Allowance, Examiner David C. THOMAS, mail date: 2/28/2020. 6 pages.	
/D.C.T/	C5	USPTO, Application No. 16/411,069, Notice of Allowance, Examiner David C. THOMAS, mail date: 3/6/2020. 6 pages.	
/D.C.T/	C6	USPTO, Application No. 16/514,931, Non-Final Office Action, Examiner David C. THOMAS, mail date: 1/29/2020. 12 pages.	
/D.C.T/	C7	USPTO, Application No. 16/514,931, Notice of Allowance, Examiner David C. THOMAS, mail date: 6/1/2020. 5 pages.	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

/D.C.T/	C8	USPTO, Application No. 16/503,382, Non-Final Office Action, Examiner Stephanie K. MUMMERT, mail date: 1/24/2020. 12 pages.	
/D.C.T/	C9	USPTO, Application No. 16/503,382, Notice of Allowance with notice of references cited, Examiner Stephanie K. MUMMERT, mail date: 7/15/2020. 12 pages.	
/D.C.T/	C10	USPTO, Application No. 16/503,382, Corrected Notice of Allowance with notice of references cited, Examiner Stephanie K. MUMMERT, mail date: 7/29/2020. 4 pages.	

EXAMINER SIGNATURE			
Examiner Signature	/DAVID C THOMAS/	Date Considered	11/09/2020

*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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2020-09-29
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1	6582908	B2	2003-06-24	FODOR et al.	
	A2	7406385	B2	2008-07-29	SORENSON	
	A3	8029993	B2	2011-10-04	MIKAWA	
	A4	9249460	B2	2016-02-02	PUSHKAREV et al.	
	A5	9260753	B2	2016-02-16	XIE et al.	
	A6	10287630	B2	2019-05-14	XIE et al.	
	A7	10287631	B2	2019-05-14	SALK 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
	A8	20040209299	A1	2004-10-21	PINTER et al.	
	A9	20070172839	A1	2007-07-26	SMITH et al.	
	A10	20070172873	A1	2007-07-26	BRENNER et al.	
	A11	20110160078	A1	2011-06-30	FODOR et al.	
	A12	20120208724	A1	2012-08-16	STEEMERS et al.	
	A13	20120220494	A1	2012-08-30	SAMUELS et al.	
	A14	20120244525	A1	2012-09-27	HENDRICKSON	
	A15	20140057799	A1	2014-02-27	JOHNSON et al.	
	A16	20140134610	A1	2014-05-15	PHAM et al.	
	A17	20140329698	A1	2014-11-06	BIGNELL et al.	
	A18	20180363051	A1	2018-12-20	SALK et al.	
	A19	20180363052	A1	2018-12-20	SALK et al.	
	A20	20180363053	A1	2018-12-20	SALK et al.	
	A21	20190093161	A1	2019-03-28	SALK et al.	
	A22	20190093162	A1	2019-03-28	SALK et al.	
	A23	20190119748	A1	2019-04-25	SALK et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	A24	20190119749	A1	2019-04-25	SALK et al.	
	A25	20190271040	A1	2019-09-05	SALK et al.	
	A26	20190284626	A1	2019-09-19	SALK et al.	
	A27	20190284627	A1	2019-09-19	SALK et al.	
	A28	20190292597	A1	2019-09-26	SALK et al.	
	A29	20190323082	A1	2019-10-24	SALK et al.	
	A30	20190338358	A1	2019-11-07	SALK et al.	
	A31	20190352714	A1	2019-11-21	SALK et al.	

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]
	B1	2012042374	WO	A2	2012-04-05	Taipale et al.		
	B2	2012129363	WO	A2	2012-09-27	Xie et al. (PRESIDENT AND FELLOWS OF HARVARD COLLEGE)		

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]
	C1	BENTLEY, David R. et al., "Accurate whole human genome sequencing using reversible terminator chemistry," Nature 456:53-59 (2008)	
	C2	CHIU, R.W.K. et al., "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma," PNAS 105:20458-20463 (2008)	
	C3	DIEHL, Frank et al., "Basic - Alimentary Tract - Analysis of Mutations in DNA Isolated From Plasma and Stool of Colorectal Cancer Patients", Gastroenterology 135:489–498 (2008)	
	C4	DIEHL, Frank et al., "Detection and quantification of mutations in the plasma of patients with colorectal tumors," PNAS 102:16368-16373 (2005)	
	C5	DING, Li et al., "Analysis of next-generation genomic data in cancer: accomplishments and challenges," Human Molecular Genetics, 19:R188-R196 (2010)	
	C6	EPO, Examination Report for EP13764186.6, dated 1/31/2017, 7 pages.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C7	EWING, B. et al. “Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities” Genome Research (1998) Vol. 8, pp. 186-194.	
	C8	FLEISCHHACKER, M. et al., "Circulating nucleic acids (CNAs) and cancer-A survey," Biochimica et Biophysica Acta. 1775:181-232 (2007)	
	C9	FONG, Siew Lee et al., "Comparison of 7 Methods for Extracting Cell-Free DNA from Serum Samples of Colorectal Cancer Patients," Clinical Chemistry, 55:587-598 (2009)	
	C10	FORSHEW, T. et al. “Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA” Science Translational Medicine (2012) Vol. 4, Issue 136, 136ra68, 14 pages.	
	C11	GORDON, David J. et al., "Causes and consequences of aneuploidy in cancer," Nature Reviews Genetics 13:189-203 (2012)	
	C12	JIANG, Hui et al., "SeqMap: mapping massive amount of oligonucleotides to the genome," Bioinformatics 24:2395-2396 (2008)	
	C13	KAO, Wei-Chun et al, "BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing," Genome Research 19:1884-1895 (2009)	
	C14	KIRCHER, Martin et al., "Improved base calling for the Illumina Genome Analyzer using machine learning strategies," Genome Biology 10:R83 (2009)	
	C15	KIRSCH, Stefan et al., "Sequence error storms and the landscape of mutations in cancer," PNAS 109:14289-14290 (2012)	
	C16	LEDERGERBER, Christian et al., "Base-calling for next-generation sequencing platforms," Briefings in Bioinformatics 12:489-497 (2011)	
	C17	LI, Heng et al., "Fast and accurate short read alignment with Burrows–Wheeler transform," Bioinformatics 25:1754-1760 (2009)	
	C18	LI, Heng et al., "Mapping short DNA sequencing reads and calling variants using mapping quality scores," Genome Research, 18:1851-1858 (2008)	
	C19	LIANG, Kuo-Ching et al., "Bayesian Basecalling for DNA Sequence Analysis Using Hidden Markov Models," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4:430-440 (2007)	
	C20	LIAO, Gary J.W. et al., "Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles," Clinical Chemistry 57:92-101 (2011)	
	C21	LO, Y.M. Dennis et al., "Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis," Am. J. Hum. Genet. 62:768-775 (1998)	
	C22	LUNTER, Gerton et al., "Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads," Genome Research 21:936-939 (2011)	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/ 0524

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C23	McKERNAN, Kevin Judd et al., "Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding," Genome Research 19:1527-1541 (2009)	
	C24	MERTES, Florian et al., "Targeted enrichment of genomic DNA regions for next-generation sequencing," Briefings in Functional Genomics, 10:374-386 (2011)	
	C25	MEYER, M. et al. "Targeted high-throughput sequencing of tagged nucleic acid samples" Nucleic Acids Research (2007) Vol. 35, No. 15, e97, 5 pages.	
	C26	MEYERSON, Matthew et al., "Advances in understanding cancer genomes through second-generation sequencing," Nature Reviews Genetics, 11:685-696 (2010)	
	C27	NARAYAN, A. et al. "Ultrasensitive Measurement of Hotspot Mutations in Tumor DNA in Blood Using Error-Suppressed Multiplexed Deep Sequencing" Cancer Research (2012) Vol. 72, No. 14, pp. 3492-3498.	
	C28	NIELSEN, Rasmus et al., "Genotype and SNP calling from next-generation sequencing data," Nature Reviews Genetics, 12:443-451 (2011)	
	C29	QUINLAN, Aaron R. et al., "Pyrobayes: an improved base caller for SNP discovery in pyrosequences," Nature Methods, 5:179-181 (2008)	
	C30	REDON, Richard et al., "Global variation in copy number in the human genome," Nature 444:444-454 (2006)	
	C31	RIZZO, J. et al. "Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing" Cancer Prevention Research (2012), Vol. 5, No. 7, pp. 887-900.	
	C32	SCHMITT, Michael W. et al., "Detection of ultra-rare mutations by next-generation sequencing," PNAS 109:14508-14513 (2012)	
	C33	SCHWARZENBACH, H. et al., "Cell-free nucleic acids as biomarkers in cancer patients," Nature Reviews, 11, 426-437, 2011	
	C34	SCHWEIGER, Michal R. et al., "Genome-Wide Massively Parallel Sequencing of Formaldehyde Fixed-Paraffin Embedded (FFPE) Tumor Tissues for Copy-Number- and Mutation-Analysis," PLOS One 4:e5548 (2009)	
	C35	SEHNERT, Amy J. et al., "Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood," Clinical Chemistry 57:1042-1049 (2011)	
	C36	SPARKS, Andrew B., et al., "Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy," Prenatal Diagnosis, 32:3-9 (2012)	
	C37	TEER, Jamie K. et al, "Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding," Genome Research 20:1420-1431 (2010)	
	C38	USPTO, Application No. 16/120,019, Notice of Allowance, Examiner David C. THOMAS, mailed 10/30/2019. 15 pgs.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/ 0525

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

	C39	USPTO, Application No. 16/120,072, Notice of Allowance, Examiner Stephanie Kane MUMMERT, mailed 6/26/2019. 16 pages.	
	C40	USPTO, Application No. 16/120,091, Second Notice of Allowance, Examiner David C. THOMAS, mailed 6/10/2019. 8 pages.	
	C41	USPTO, Application No. 16/411,045, Non-Final Office Action, Examiner David C. THOMAS, mailed 7/25/2019. 20 pages.	
	C42	USPTO, Application No. 16/411,045, Notice of Allowance, Examiner David C. THOMAS, mail date: 11/15/2019. 16 pgs.	
	C43	USPTO, Application No. 16/411,068, Non-Final Office Action, Examiner David C. THOMAS, mail date: 11/20/2019. 28 pgs.	
	C44	USPTO, Application No. 16/411,069, Non-Final Office Action, Examiner David C. THOMAS, mail date: 11/26/2019. 28 pgs.	
	C45	WAGLE, Nikhil et al., "High-Throughput Detection of Actionable Genomic Alterations in Clinical Tumor Samples by Targeted, Massively Parallel Sequencing," AACR Cancer Discovery 2:82-93 (2011)	

EXAMINER SIGNATURE			
Examiner Signature	/DAVID C THOMAS/	Date Considered	11/09/2020

*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. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2019-12-16
Name/Print	Lara J. Dueppen	Registration Number	65,002

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. /D.C.T/

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	72227-8043.US07
Confirmation Number	1114

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
	A1	5308751	A	1994-05-03	Ohkawa et al.	
	A2	6251610	B1	2001-06-26	Gupte et al.	
	A3	6958225	B2	2005-10-25	Dong	
	A4	7214490	B2	2007-05-08	Su et al.	
	A5	7267966	B2	2007-09-11	Dong et al.	
	A6	7297778	B2	2007-11-20	Matsuzaki et al.	
	A7	7452699	B2	2008-11-18	Makrigiorgos et al.	
	A8	7459273	B2	2008-12-02	Jones et al.	
	A9	7741463	B2	2010-06-22	Gormley et al.	
	A10	8148068	B2	2012-04-03	Brenner et al.	
	A11	8153375	B2	2012-04-10	Travers, Kevin et al.	
	A12	8715967	B2	2014-05-06	Casbon, James et al.	
	A13	8741606	B2	2014-06-03	Casbon et al.	
	A14	9080210	B2	2015-07-14	Van Eijk, Michael J., et al.	
	A15	9476095	B2	2016-10-25	Vogelstein, Bert et al.	
	A16	9745627	B2	2017-08-29	Van Eijk, Michael J., et al.	
	A17	9862995	B2	2018-01-09	Patel, Abhijit Ajit	
	A18	9898577	B2	2018-02-20	Van Eijk, Michael J., et al.	
	A19	10011871	B2	2018-07-03	Bielas, Jason H.	
	A20	10023907	B2	2018-07-17	Van Eijk, Michael J., et al.	
	A21	10202646	B2	2019-02-12	Fodor, Stephen, 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
	A22	20070128624	A1	2007-06-07	Gormley et al.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	A23	20080167195	A1	2008-07-10	Li et al.	
	A24	20090298075	A1	2009-12-03	Travers et al.	
	A25	20100069263	A1	2010-03-18	Shendure et al.	
	A26	20100222238	A1	2010-09-02	Smith et al.	
	A27	20110301042	A1	2011-12-08	Steinmann et al.	
	A28	20120165202	A1	2012-06-28	Porreca, Gregory et al.	
	A29	20120238738	A1	2012-09-20	Hendrickson et al.	
	A30	20150024950	A1	2015-01-22	Bielas et al.	
	A31	20160319345	A1	2016-11-03	Gnerre et al.	
	A32	20170247687	A1	2017-08-31	Shendure et al.	
	A33	20180023135	A1	2018-01-25	Van Eijk, Michael J., et al.	
	A34	20180363048	A1	2018-12-20	Bielas, Jason H.	
	A35	20180363049	A1	2018-12-20	Bielas, Jason H.	

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]
	B1 **	2006113422	WO	A2	2006-10-26	Ud Gov Health & Human Serv		
	B2 **	2011021102	WO	A2	2011-02-24	POPULATION GENETICS TECHNOLOGIES LTD.		
	B3 **	2012061832	WO	A1	2012-05-10	ILLUMINA, INC.		

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]
	C1 **	Ameur A, Stewart JB, Freyer C, Hagstrom E, Ingman M, Larsson N-G, et al. UltraDeep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. PLoS Genet. 2011;7:e1 002028.	
	C2 **	Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch M, D'Ascenzo M, et al. Whole exome capture in solution with 3 Gbp of data. Genome Bioi. 2010;11: R62:1-8.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C3 **	Boyd SO, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. Science Translational Medicine. 2009;1 :12ra23-12ra23.	
	C4 **	Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci USA. 2008; 105:13081-6.	
	C5 **	Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, et al. Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat Methods. 2012;9:78-80.	
	C6 **	Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. Nucleic Acids Research. 2011 ;39:e81-e.	
	C7 **	Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. Embryonic stem cells and somatic cells differ in mutation frequency and type. Proc Natl Acad Sci USA. 2002;99:3586-90.	
	C8 **	Chiu RWK, Akolekar R, Zheng YWL, Leung TY, Sun H, Chan KCA, et al. Noninvasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011 ;342:c7401.	
	C9 **	De Grassi A, Segala C, Iannelli F, Volorio S, Bertario L, Radice P, et al. Ultradeep Sequencing of a Human Ultraconserved Region Reveals Somatic and Constitutional Genomic Instability. PLoS Biol. 2010;8:e1000275.	
	C10 **	Ding L, Ley T J, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 21 05;481 :506-9.	
	C11 **	Druley TE, Vallania FLM, Wegner OJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA, Nat Methods. 2009;6 :263-5.	
	C12 **	Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obsbet Gynecol. 2011;204:205e1-11.	
	C13 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated October 4, 2018, 6 pages	
	C14 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated June 8, 2018, 5 pages	
	C15 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated August 10, 2017, 7 pages	
	C16 **	EUROPEAN PATENT OFFICE, Examination Report for EP13764186.6, dated May 13, 2016, 6 pages	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C17 **	EUROPEAN PATENT OFFICE, Extended European Search Report for EP13764186.6, dated September 8, 2015, 3 pages	
	C18 **	Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake Sr. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA. 2008;1 05:16266-71.	
	C19 **	Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, et al. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic Acids Research. 2012;40:e2-e.	
	C20 **	Fordyce SL, Avila-Areas MC, Rockenbauer E, B0rsting C, Frank-Hansen R, Petersen FT, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. BioTechniques. 2011;51 :127-33.	
	C21 **	Fu GK, Hu J, Wang P-H, Fodor SPA. Counting individual DNA molecules by the stochastic attachment of diverse labels. Proc Natl Acad Sci USA. 2011;108:9026-31.	
	C22 **	Garcia-Garcera M, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, et al. Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. PLoS ONE. 2011;6:e24161.	
	C23 **	Greaves LC, et al. (2009) Quantification of mitochondrial DNA mutation load. Aging Cell 8:566-572.	
	C24 **	Haag-Liautard C, et al. (2008) Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Bioi 6:e204.	
	C25 **	He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464:610-4.	
	C26 **	Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501-509.	
	C27 **	Hyman RW, Herndon CN, Jiang H, Palm C, Fukushima M, Bernstein D, et al. The dynamics of the vaginal microbiome during infertility therapy with in vitro fertilization embryo transfer. J Assist Reprod Genet. 2012;29:105-15.	
	C28 **	Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA. 2011;108:20166-71.	
	C29 **	Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc Natl Acad Sci USA 93:12382-12387.	
	C30 **	Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng. 2003;96:317-23.	
	C31 **	Kasai H, et al.(1993) Formation, inhibition of formation, and repair of oxidative 8-hydroxyguanine DNA damage. Basic Life Sci 61:257-262.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C32 **	Kaur M, Makrigiorgos GM. Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. Nucleic Acids Res. 2003;31:26e1-7.	
	C33 **	Kennedy SR, Loeb LA, Herr AJ (2012) Somatic mutations in aging, cancer and neurodegeneration. Mech Ageing Dev 133(4):118-26.	
	C34 **	Khaidakov M, Heflich RH, Manjanatha MG, Myers MB, Aidoo A (2003) Accumulation of point mutations in mitochondrial DNA of aging mice. Mutat Res 526:1-7.	
	C35 **	Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci USA. 2011;108:9530-5.	
	C36 **	KIVIOJA et al., "Counting absolute numbers of molecules using unique molecular identifiers," Nature Methods 9, dated January 1, 2012, pages 72-74	
	C37 **	Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner OJ. Amplification free Illumine sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods. 2009; 6:291-5.	
	C38 **	Kraytsberg Y, Nicholas A, Caro P, Khrapko K (2008) Single molecule PCR in mtDNA mutational analysis: Genuine mutations vs. damage bypass-derived artifacts. Methods 46:269-273.	
	C39 **	Kunkel, TA. Mutational specificity of depurination. Proc Natl Acad Sci USA. 1984; 81:1494-98.	
	C40 **	LaTuga MS, Ellis JC, Cotton CM, Goldberg RN, Wynn JL, Jackson RB, et al. Beyond bacteria: a study of the enteric microbial consortium in extremely low birth weight infants. PLoS ONE. 2011; 6:e27858.	
	C41 **	Lecroq B, Lejzerowicz F, Bechar D, Christen R, Esling P, Baerlocher L, et al. Ultradeep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc Natl Acad Sci USA. 2011;108:13177-82.	
	C42 **	Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF (2002) High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Hum Mol Genet 11:133-145.	
	C43 **	Lindahl T, Wood RD. Quality control by DNA repair. Science. 1999;286:1897-1905.	
	C44 **	Lynch AM, Sasaki jC, Elespuru R, Jacobson-Kram D, Thybaud V, et al. New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen. 2011;52(3):205-23.	
	C45 **	Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature. 2011 ;480:368-71.	
	C46 **	McBride T J, Preston BD, Loeb LA (1991) Mutagenic spectrum resulting from DNA damage by oxygen radicals. Biochemistry 30:207-213.	
	C47 **	McCloskey ML, Stager R, Hansen RS, Laird CD. Encoding PCR products with batch-stamps and barcodes. Biochem Genet. 2007; 45:761-7.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C48 **	Metzker ML. Sequencing technologies--the next generation. Nat Rev Genet. 2010;11:31-46.	
	C49 **	Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. Nucleic Acids Research. 1990;18:1687-91.	
	C50 **	Miner BE, Stager RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Research. 2004;32:e135.	
	C51 **	Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GO, et al. The human gut virome: interindividual variation and dynamic response to diet. Genome Res. 2011; 21:1616-25.	
	C52 **	Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008; 105:10513-8.	
	C53 **	Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS ONE. 2011;6:e24907.	
	C54 **	Out AA, van Minderhout IJHM, Goeman JJ, Ariyurek Y, Ossowski S, Schneeberger K, et al. Deep sequencing to reveal new variants in pooled DNA samples. Hum Mutat. 2009;30:1703-12.	
	C55 **	Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., Mcinerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature 461, 814-818.	
	C56 **	Parsons T J, et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363-368.	
	C57 **	Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 2008;5:1005-10.	
	C58 **	ROBERTS et al., "Short template amplicon and multiplex megaprimer - enabled relay (STAMMER) sequencing, a simultaneous approach to higher throughput sequence-based typing of polymorphic genes", Immunogenetics (2010), Vol. 62, 253-60	
	C59 **	Salk J, Fox E, Loeb L. Mutational heterogeneity in human cancers: origin and consequences. Annual Review of Pathology. 2009;5:51-75.	
	C60 **	Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20:273-80.	
	C61 **	Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135-45.	
	C62 **	Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature. 1991;349:431-4.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C63 **	SHIROGUCHI et al., “Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes", PNAS 109, dated January 9, 2012, pages 1347-1652	
	C64 **	Song S, et al. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. Proc Natl Acad Sci USA 102:4990-4995.	
	C65 **	Stiller M, Green RE, Ronan M, Simons JF, Du L, HeW., et al. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Aced Sci USA. 2006; 103:13578-84.	
	C66 **	Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029-1032.	
	C67 **	Thomas DC, Roberts JD, Sabatino RD, Myers TW, et al. Fidelity of mammalian DNA replication and replicative DNA polymerases. Biochemistry. 1991;30:11751-9.	
	C68 **	Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 2010; 38:159e1-8.	
	C69 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Search Report and Written Opinion for PCT/US2013/032665, dated July 9, 2013, 14 pages	
	C70 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Non-Final Office Action for U.S. Patent Application 16/120,072 mailed December 14, 2018. 16 pages.	
	C71 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Non-Final Office Action for U.S. Patent Application No. 16/120,019. Mail Date: November 19, 2018. 16 pages.	
	C72	**UNITED STATES PATENT AND TRADEMARK OFFICE, Notice of Allowance for U.S. Patent Application No. 15/660,785. Mail Date: December 27, 2018. 8 pages.**	
	C73 **	UNITED STATES PATENT AND TRADEMARK OFFICE, Restriction Requirement for U.S. Patent Application No. 16/120,091. Mail Date: November 23, 2018. 7 pages.	
	C74	**UNITED STATES PATENT AND TRADEMARK OFFICE, Notice of Allowance for U.S. Patent Application No. 16/120,091. Mail Date: March 4, 2019. 9 pages.**	
	C75	**UNITED STATES PATENT AND TRADEMARK OFFICE, Final Office Action for U.S. Patent Application No. 16/120,019. Mail Date: June 7, 2019. 17 pages.**	
	C76 **	Vandenbroucke I, Van Marek H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques. 2011;51 :167-77.	
	C77 **	Vermulst M, et al. (2007) Mitochondrial point mutations do not limit the natural lifespan of mice. Nat Genet 39:540-543.	

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

	C78 **	Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17:1195-201.	
	C79 **	Wiemann S, Stegemann J, Grothues D, Bosch A, Estivill X, Schwager C, Zimmermann J, Voss H, Ansorge W. (1995) Simultaneous On-Line DNA Sequencing on Both Strands with Two Fluorescent Dyes. Analytical Biochemistry, Vol. 224, pp. 117-121.	
	C80 **	Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbial. 2011 ;49:3463-9.	
	C81 **	Zagordi O, Klein R, Daumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Research. 2010;38:7400-9.	

EXAMINER SIGNATURE			
Examiner Signature	/DAVID C THOMAS/	Date Considered	11/09/2020

*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. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	72227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

REFERENCES INCLUDED

Copies of non-patent references C72, C74, and C75 are included. Copies of the foreign and non-patent references marked with a double asterisk are not included, because they were previously cited by or submitted to the Office in prior application number 15/660,785, filed July 26, 2017, or application number 14/386,800, filed September 20, 2014, and relied upon in this application for an earlier filing date under 35 U.S.C. §120. The instant application number 16/118,306 is a continuation of application number 15/660,785.

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2018-06-21
Name/Print	Lara J. Dueppen	Registration Number	65,002

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. /D.C.T/

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	1	("20150024950").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/12 18:21
S1	1	("10287631").PN.	USPAT; USOCR	OR	OFF	2020/11/09 16:33
S2	1	("9752188").PN.	USPAT; USOCR	OR	OFF	2020/11/09 16:36
S3	40,249	sequence read$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:24
S4	0	S3 and amplif$7 each strand	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:25
S5	0	S3 and amplif$7 each original strand	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S6	19,195	S3 and amplif$7	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S7	4,281	S6 and consensus sequence	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S8	1,361	S7 and both strands	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:27
S9	1,283	S8 and double-stranded	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:28
S10	83	S9 and y-shaped adapt$2	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:28
S11	121	S8 and double-stranded sequencing	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:31
S12	513	c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:48
S13	183	c12q 1/6876	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49

S14	613	c12q 1/6869	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S15	84	S14 and c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S16	21	S13 and c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S17	27	S14 and c12q 1/6876	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:51
S18	1	("9970054").PN.	USPAT; USOCR	OR	OFF	2020/11/10 11:58
S19	1	("20190093161").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/10 11:59
S20	1	("20090298075").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/10 12:17

11/12/2020 6:31:05 PM
C:\Users\dthomas1\Documents\EAST\Workspaces\16118306.wsp

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
/D.C.T/	A1	2014/0155274	A1	2014-06-05	Xie et al.	

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]
/D.C.T/	C1	EPO, Extended European Search Report for Application No. EP20166917.3, dated September 30, 2020. 6 pages.	
/D.C.T/	C2	USPTO, Application No. 16/908,611, Non-Final Office Action, Examiner David C. THOMAS, mail date: 10/9/2020. 28 pages.	

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	11/09/2020

*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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

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

☒ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2020-10-24
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	JESSE SALK, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1637
FILED:	AUGUST 30, 2018	EXAMINER:	DAVID C. THOMAS
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

AMENDMENT AND RESPONSE

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

Dear Commissioner:

In response to the Non-Final Office Action mailed November 18, 2020, please amend the above-identified U.S. patent application as follows:

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

Remarks begin on page 8 of this paper.

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1	20120058468	A1	2012-03-08	McKeown	
	A2	20140329282	A1	2015-01-22	Nelson et al.	

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]
	C1	USPTO, Application No. 16/118,290, Non-Final Office Action, Examiner David C. THOMAS, mail date: 12/14/2020. 33 pages.	
	C2	USPTO, Application No. 17/008,395, Non-Final Office Action, Examiner Stephanie Kane MUMMERT, mail date: 01/07/2021, 20 pages	
	C3	USPTO, Application No. 16/118,305, Non-Final Office Action, Examiner David C. THOMAS, mail date: 1/13/2021	
	C4	USPTO, Application No. 16/908,611, Notice of Allowance, Examiner David C. THOMAS, mail date: 1/29/2021. 11 pages.	
	C5	USPTO, Application No. 16/119,471, Non-Final Office Action, Examiner David C THOMAS, mail date: 02/02/2021, 32 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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-02-18
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Electronic Patent Application Fee Transmittal

Application Number:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Filer:	Lara J. Dueppen/Jonathan Petrovich
Attorney Docket Number:	072227-8043.US07

Filed as Small 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($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	130	130
		Total in USD ($)		**130**

Electronic Acknowledgement Receipt

EFS ID:	41962520
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	18-FEB-2021
Filing Date:	30-AUG-2018
Time Stamp:	19:24:07
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$130
RAM confirmation Number	E20212HJ24564982
Deposit Account	500665
Authorized User	Jonathan Petrovich

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

37 CFR 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		2021-02-18-NFOA-Response-072227-8043US07.pdf	168442 de277c6e611999556ef278cb01b55a97922b44ec	yes	14

Multipart Description/PDF files in .zip description

Document Description	Start	End
Applicant Arguments/Remarks Made in an Amendment	8	14
Claims	2	7
Amendment/Req. Reconsideration-After Non-Final Reject	1	1

Warnings:

Information:

2	Information Disclosure Statement (IDS) Form (SB08)	2021-02-18-Supplemental-IDS-072227-8043US07.pdf	69569 7e149de1ae7a75051b558ae78f3cc6b02a1089c7	no	2

Warnings:

Information:

This is not an USPTO supplied IDS fillable form

3	Non Patent Literature	C1-2020-12-14-NFOA-US16118290-8043US05.PDF	3795838 e5011884eae91a3f095bdb4105292f89ed19f44c	no	33

Warnings:

Information:

4	Non Patent Literature	C2-2021-01-07-NFOA-US17008395-8043US20.PDF	2255119 f8102b5f44182454d84979ff288c3ce8d01246c2	no	20

Warnings:

Information:

5	Non Patent Literature	C3-2021-01-13-NFOA-US16118305-8043US06.PDF	3241325 2a2cded69a26473588595b79de77d8e38195498c	no	33
Warnings:					
Information:					
6	Non Patent Literature	C4-2021-01-29-NOA-US16908611-8043US19.PDF	1154916 464feb54839d9e8f4cc8bd77544ba97053ba1112	no	11
Warnings:					
Information:					
7	Non Patent Literature	C5-2021-02-02-NFOA-US16119471-8043US08.PDF	3151124 6520cf912723552411efc4453a6d4af6141321d3	no	32
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	30835 3e84e52ef9dec8c736e024ce26a8f22ba5baafa8a	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	13867168	

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

Claims 37-66 were pending when the present Office Action was mailed on November 18, 2020, with claim 64 withdrawn from consideration. In this response, claims 55, 58, 59, 63 and 64 have been amended to correct minor grammatic errors or to clarify certain features of these claims; the foregoing amendments are made without prejudice to pursuing these claims in unamended or other forms in a continuation or other application. Claims 46-54 have been canceled in this response without conceding the merits of the rejection of these claims and without prejudice to pursuing these claims in a continuation application. New claim 67 has been added with support in the original specification. Because no new matter is introduced, Applicant respectfully requests entry of this amendment. Upon entry, claims 37-45 and 55-66 will be pending and under examination, with claim 64 withdrawn from consideration.

Double Patenting

Claims 37-63, 65 and 66 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-23 of U.S. Patent No. 10,287,631.

Claims 37-63, 65 and 66 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/118,290.

Claims 37-63, 65 and 66 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/503,398.

Applicant requests that the Office continue to hold the above listed double-patenting rejections to the issued patent and copending patent applications in abeyance

until at least one pending claim of the instant application is found to be otherwise allowable.

Claim Objections

Claims 55 and 59 are objected to because of informalities. In this response, Applicant has amended claims 55 and 59 in the manner suggested by the Examiner.

Claim Rejections - 35 USC § 112

Claim 58 is rejected under 35 U.S.C. 112 (pre-AIA), second paragraph, as allegedly being indefinite for failing to particularly point out and distinctly claim the subject matter which the inventor or a joint inventor (or for applications subject to pre-AIA 35 U.S.C. 112, the applicant), regards as the invention. More particularly, the Office Action states that "the phrase 'and preferably' renders the claim indefinite." (OA at page 10) Without commenting on or conceding the merits of this rejection, and in the interest of advancing prosecution, Applicant has amended claim 58 to remove the phrase 'and preferably'. As such, Applicant respectfully requests that the Section 112 rejection of claim 58 be withdrawn.

Claim Rejections - 35 USC § 102

Claim(s) 37-39, 41, 42, 45-48, 52-58 and 60-63 is/are rejected under pre-AIA 35 U.S.C. 102(b) as allegedly being anticipated by Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019). In this response, claims 46-48 and 52-54 have been cancelled and the rejection of these claims under Section 102 is therefore moot. With respect to claims 37-39, 41, 42, 45, 55-58 and 60-63, and for at least the reasons explained below, Applicant respectfully traverses this rejection.

Claim 37 is patentable over Travers because Travers fails to disclose or suggest each and every feature of claim 37 including, *inter alia*, "amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand

generating a distinct yet related set of amplified target nucleic acid products;…" The Office Action alleges that Travers discloses this feature by teaching that "template constructs may be prepared from PCR products to provide a population of identical molecules from which to generate sequence reads" (November 18th Office Action at page 11, citing Travers at ¶¶[0054], [0066] and [0088]). Applicant respectfully disagrees.

In contrast to the Office Action's interpretation, Travers does not teach any steps that would result in distinct sets of amplified target nucleic acid products (e.g., a distinct set from each original strand). Instead, Travers describes method steps for use in single molecule sequencing reactions that includes: *linking* sense and antisense strands of a double-stranded nucleic acid molecule together with a "linking segment" (e.g., using exogenous hairpin adapters (see, e.g., Travers ¶[0051])) and amplification of these linked original strands "to ensure complete, and preferably redundant sequencing of the entire template" (¶[0081]). By linking the sense and antisense strands together, Travers ensures that only a single set of amplified products are generated from any one particular double-stranded molecule, with every member of the set being an identical sequence. As stated in cited paragraph [0054], Travers discloses that each template molecule includes sequence segments corresponding to (1) the sense strand, (2) the antisense strand and (3) the linking segment(s) of the hairpin adapter(s), such that "a *single template molecule*…can be sequenced in one integrated process" (emphasis added).

Nowhere does Travers disclose or suggest separating the sense strand from the antisense strand of a double-stranded DNA fragment in a manner that such original strand amplification products are distinct yet relatable. Moreover, one of ordinary skill in the art would understand from a reading of Travers that separating the sense and antisense strands *before*, *during* or *after* amplification of the template molecules would thwart the entire purpose of Travers (e.g., to generate a consensus sequence of the double stranded molecule by redundant sequencing of the entire single template). As Travers is silent about "amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target

nucleic acid products," as recited in claim 37, Travers also does not disclose how one of ordinary skill in the art would perform additional steps recited in claim 37, including, *inter alia*:

> confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands;
>
> comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand; and
>
> generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary.

Applicant further submits that a person skilled in the art would not be motivated to modify Travers to include these recited features.

Accordingly, for at least the foregoing reasons, Travers fails to disclose or suggest all of the features of claim 37. Therefore, the Section 102 rejection of claim 37 should be withdrawn.

Independent claim 55, as amended, recites several features similar to the features of claim 37. For example, claim 55 recites, *inter alia*, "generating a first set of duplicates of an original forward strand of the tagged DNA molecule and a second set of duplicates of an original complement strand of the tagged DNA molecule;..." As discussed above, Travers does not disclose or suggest generating more than one set of duplicates from the

original strands (e.g., the original forward strand and the original complement strand). As such, Travers does not disclose or suggest how one skilled in the art would perform any following steps recited in claim 55. Accordingly, for at least the foregoing reasons, Travers fails to disclose or suggest all of the features of claim 55, and the Section 102 rejection of 55 should be withdrawn.

Claims 38, 39, 41, 42, and 45, depend from base claim 37 and claims 56-58 and 60-63 depend from base claim 55. Accordingly, the Section 102 rejection of claims 38, 39, 41, 42, 45, 56-58 and 60-63 should be withdrawn for at least the reasons discussed above with respect to claims 37 and 55, and for the additional features of these dependent claims.

Claim Rejections - 35 USC § 103 (Travers and Xie)

Claims 40 and 59 is/are rejected under pre-AIA 35 U.S.C. 103(a) as allegedly being unpatentable over Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019) in view of Xie et al. U.S. Patent Pub. No. 2014/0155274, cited on IDS of 10/26/2020).

Claims 40 and 59 depend from independent claims 37 and 55, respectively. Travers fails to disclose the elements recited in claims 37 and 55 for at least the reasons discussed in detail above. Without commenting on or conceding that Xie discloses the elements for which this reference was cited, Xie nevertheless fails to supply the deficiencies of Travers. Therefore claims 37 and 55 are patentable over Travers in view of Xie. Accordingly, dependent claims 40 and 59 are allowable over the combination of Travers and Xie for at least the reason that these references, either alone or in combination, fail to disclose or suggest all of the features of independent claims 37 and 55, and for the additional features of these dependent claims. Therefore, the Section 103 rejection of claims 40 and 59 should be withdrawn and claims 40 and 59 should be allowed.

Claim Rejections - 35 USC § 103 (Travers and Bielas)

Claims 43, 44, 49-51, 65 and 66 is/are rejected under pre-AIA 35 U.S.C. 103(a) as allegedly being unpatentable over Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019) in view of Bielas et al. (U.S. Patent Pub. No. 2015/0024950, cited on IDS of 6/21/2019). In this response, claims 49-51 have been cancelled and the rejection of these claims under Section 103 is therefore moot. Remaining claims 43 and 44 depend from independent claim 37 and claims 65 and 66 depend from independent claim 55. Travers fails to disclose the elements recited in claims 37 and 55 for at least the reasons discussed in detail above. Without commenting on or conceding that Bielas discloses the elements for which this reference was cited, Bielas nevertheless fails to supply the deficiencies of Travers. Therefore claims 37 and 55 are patentable over Travers in view of Bielas. Accordingly, dependent claims 43, 44, 65 and 66 are allowable over the combination of Travers and Bielas for at least the reason that these references, either alone or in combination, fail to disclose or suggest all of the features of independent claims 37 and 55, and for the additional features of these dependent claims. Therefore, the Section 103 rejection of claims 43, 44, 65 and 66 should be withdrawn and claims 43, 44, 65 and 66 should be allowed.

New Claim

New claim 67 has been added in this response. The subject matter of this claim is supported throughout the original application as-filed, does not add any new matter to the application and is fully supported under 35 U.S.C. § 112, first paragraph. New claim 67 depends from claim 55, and is therefore patentable over the applied reference Travers, with or without Xie and/or Bielas, for at least the reasons discussed above with respect to independent claim 55, and the additional features of this new claim.

CONCLUSION

In view of the foregoing, Applicant requests reconsideration of the application and respectfully submits that the application is in condition for allowance. If the Examiner has any questions or believes a telephone conference would expedite prosecution of this application, the Examiner is encouraged to contact the undersigned at (310) 788-3349. Please charge any underpayment or credit any overpayment to our Deposit Account No. 50-0665.

Applicant reserves the right to later pursue any previously pending or other broader or narrow claims that capture any subject matter supported by the present disclosure, including subject matter that might be found disclaimed herein or by any prior prosecution. Accordingly, reviewers of this or any parent, child or related prosecution history shall not reasonably infer that the applicant has made any disclaimers or disavowals of any subject matter supported by the present application.

Dated: February 18, 2021

Respectfully submitted,

Customer No. 94991
UW/Perkins Coie LLP
Patent - LA
P.O. Box 1247
Seattle, WA 98111-1247
Phone: (310) 788-3349
Fax: (206) 332-7198

PERKINS COIE LLP

By: /Lara J. Dueppen/
Lara J. Dueppen
Registration No. 65,002

IN THE CLAIMS

The following complete listing of claims replaces all previous listings of claims in the application.

1-36. (Canceled)

37. (Previously Presented) A method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising:

amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products;

sequencing the amplified target nucleic acid products generated from each original strand;

confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands;

comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand; and

generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary.

38. (Previously Presented) The method of claim 37, wherein prior to amplifying each strand, the method further comprises ligating the double-stranded target nucleic acid molecule to an adapter molecule resulting in an asymmetric adapter-target nucleic acid complex such that each original strand generates a distinct yet related set of amplified target nucleic acid products during the amplification step.

39. (Previously Presented) The method of claim 38, wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

40. (Previously Presented) The method of claim 39, wherein the adapter molecule further comprises a nucleotide sequence that uniquely tags each original strand of the double-stranded target nucleic acid molecule such that each original strand is distinctly identifiable relative to its original complementary strand.

41. (Previously Presented) The method of claim 37, wherein the confirming step includes (i) grouping the sequence reads of the amplified target nucleic acid products into families of paired target nucleic acid strand sequences using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the double-stranded target nucleic acid molecule as an identifier, and (ii) separating each individual family of paired target nucleic acid strand sequences into a set of sequence reads from one original strand of the double-stranded target nucleic acid molecule and a set of sequence reads from the other original strand of the double stranded target nucleic acid molecule.

42. (Previously Presented) The method of claim 37, wherein the confirming step includes confirming the presence of at least about 3 sequence reads of an amplified target nucleic acid product generated from each of the original strands, and wherein the comparing step includes comparing a first consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from one original strand with a second consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from the other original strand.

43. (Previously Presented) The method of claim 37, further comprising identifying a mutation occurring at a particular position in the consensus sequence as a true mutation.

44. (Previously Presented) The method of claim 43, further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation.

45. (Previously Presented) The method of claim 37, wherein generating a consensus sequence of the double-stranded target nucleic acid molecule further comprises identifying one or more nucleotide positions that disagree between the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand, and discounting nucleotide positions in the consensus sequence that are in disagreement to provide an error-corrected sequence read of the double-stranded target nucleic acid molecule.

46-54. (Cancelled)

55. (Currently Amended) A method of generating an error-corrected consensus sequence comprising:

ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules;

for each of one or more tagged DNA molecules

generating a first set of duplicates of an original forward strand of the tagged DNA molecule and a second set of duplicates of an original complement strand of the tagged DNA molecule;

creating a single strand consensus sequence (SSCS) from the first set of duplicates of the original forward strand and a single strand

consensus sequence (SSCS) from the second set of duplicates of the original complement strand;

comparing the SSCS of the original forward strand to the SSCS of the original complement strand; and

generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward strand and the SSCS of the original complement strand are ~~complimentary~~ complementary.

56. (Previously Presented) The method of claim 55, wherein one or more adapter molecules are single molecule identifier (SMI) adapter molecules comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

57. (Previously Presented) The method of claim 56, wherein the SMI adapter molecules comprise a single-stranded SMI sequence or a double-stranded SMI sequence.

58. (Currently Amended) The method of claim 56, wherein the tagged DNA molecule comprises an SMI sequence in each of its original forward strand and original complement strand~~, and preferably wherein the SMI sequence in one strand comprises a first degenerate or semi-degenerate sequence and in the other strand comprises a second degenerate or semi-degenerate sequence that is complementary to the first such sequence~~.

59. (Currently Amended) The method of claim 55, wherein the adapter molecules comprise[[s]] a Y-shape, a U-shape, or a combination thereof.

60. (Previously Presented) The method of claim 55, wherein the tagged DNA molecule comprises an adapter molecule at each end.

61. (Previously Presented) The method of claim 55, wherein the comparing step includes pairing the SSCS of the original forward strand to the SSCS of the original complement strand using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of a particular duplex DNA molecule as an identifier.

62. (Previously Presented) The method of claim 55, further comprising sequencing one or more of the duplicates of the original forward strand and one or more of the duplicates of the original complement strand of the tagged DNA molecules using a next-generation sequencing procedure.

63. (Currently Amended) The method of claim 55, wherein the first set of duplicates of the original forward strand and the second set of duplicates of the original complement strand includes a minimum of 1 member in each set.

64. (Currently Amended and Withdrawn) The method of claim 55, wherein the first set of duplicates of the original forward strand and the second set of duplicates of the original complement strand includes a minimum of 3 members in each set.

65. (Previously Presented) The method of claim 55, wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject, and wherein the method further comprises:

comparing the error-corrected consensus sequence to a respective reference sequence; and

detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

66. (Previously Presented) The method of claim 65, further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence

variations present in the error-corrected consensus sequence and not present in the reference sequence.

67. (New) The method of claim 58 wherein the SMI sequence in one strand comprises a first degenerate or semi-degenerate sequence and in the other strand comprises a second degenerate or semi-degenerate sequence that is complementary to the first such sequence.

Electronic Patent Application Fee Transmittal

Application Number:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Filer:	Lara J. Dueppen/Jonathan Petrovich
Attorney Docket Number:	072227-8043.US07

Filed as Small 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($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	130	130
		Total in USD ($)		**130**

Electronic Acknowledgement Receipt

EFS ID:	41962679
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	18-FEB-2021
Filing Date:	30-AUG-2018
Time Stamp:	19:47:07
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$130
RAM confirmation Number	E20212HJ47340198
Deposit Account	500665
Authorized User	Jonathan Petrovich

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

37 CFR 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Fee Worksheet (SB06)	fee-info.pdf	30835 08a3c03e6878dd39ae95aad56db430c734deff05	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	30835		

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 16/118,306	Filing Date 08/30/2018	☐ 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 $50 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $230 =	
☐ 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	02/18/2021	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 22	Minus	** 30	= 0		x $50 =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	*** 3	= 0		x $240 =	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".	/LINDA HUMES/
*** 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 (02-18)
Approved for use through 11/30/2020. 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	16118306
Filing Date	2018-08-30
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	
Attorney Docket Number	072227-8043.US07

U.S.PATENTS Remove

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					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

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
	1	20030165923	A1	2003-09-04	Li et al.	
	2	20120094847	A1	2012-04-19	Warthmann et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	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]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16118306
	Filing Date	2018-08-30
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	
	Attorney Docket Number	072227-8043.US07

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, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	USPTO, Application No. 16/118,286, Non-Final Office Action, Examiner David C. THOMAS, mail date: 3/1/2021	

If you wish to add additional non-patent literature document citation information please click the Add button Add

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.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16118306
	Filing Date	2018-08-30
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	
	Attorney Docket Number	072227-8043.US07

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-03-22
Name/Print	Lara J. Dueppen	Registration Number	65002

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.**

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 the Freedom of Information Act requires disclosure of these record s.

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 inspections 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.

Electronic Patent Application Fee Transmittal

Application Number:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Filer:	Lara J. Dueppen/Jonathan Petrovich
Attorney Docket Number:	072227-8043.US07

Filed as Small 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($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	2806	1	130	130
		Total in USD ($)		**130**

Electronic Acknowledgement Receipt

EFS ID:	42247957
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Jonathan Petrovich
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	22-MAR-2021
Filing Date:	30-AUG-2018
Time Stamp:	17:30:17
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$130
RAM confirmation Number	E20213LH33010530
Deposit Account	500665
Authorized User	Jonathan Petrovich

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

37 CFR 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Information Disclosure Statement (IDS) Form (SB08)	2021-03-22-Supplemental-IDS-072227-8043US07.PDF	1034989 fbe128179f3578cc2d4c3bd90f7c6900465b8d0b	no	4
Warnings:					
Information:					
2	Non Patent Literature	C1-2021-03-01-NFOA-US16118286-8043US04.PDF	1167777 8fd053d6460f497235ad05ba1b9c2300640afab9	no	11
Warnings:					
Information:					
3	Fee Worksheet (SB06)	fee-info.pdf	30835 dbbee2f27ea2a0cfa848273c19f59177c0015cd3	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	2233601		

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.

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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

94991 7590 03/29/2021

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
03/29/2021	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):

patentprocurement@perkinscoie.com

Office Action Summary	Application No. 16/118,306	Applicant(s) Salk et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status No

-- 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 18 February 2021.

☐ 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) 37-45 and 55-67 is/are pending in the application.

5a) Of the above claim(s) 64 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 37-45,55-63 and 65-67 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 _____ 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 3/22/2021, 2/18/2021.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application is being examined under the pre-AIA first to invent provisions. Applicant's amendment to the claims filed February 18, 2021 is acknowledged. Claims 55, 58, 59 and 63 (currently amended), 37-45, 56, 57, 60-62, 65 and 66 (original), and claim 67 (newly added) will be examined on the merits. Claims 46-54 are newly canceled and claims 1-36 were previously canceled. Claim 64 was previously withdrawn and is currently amended.

Double Patenting

Nonstatutory Obviousness-type Double Patenting Rejections

2. 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. See *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); and *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) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

With respect to the non-statutory double patenting rejection(s) made in this Office action, note as follows. The use of the terminology "defined in 35 U.S.C. §154 to §156 and §173" in a terminal disclaimer can result in the terminal disclaimer being found improper. To address this, note that a proper terminal disclaimer need only disclaim the patent's remaining "full statutory term" as defined in 35 U.S.C., without specifying 35 U.S.C. 154 and 173. This is so, because the "full statutory term" inherently is a statutorily defined item.

The USPTO internet Web site contains terminal disclaimer forms which may be used. Please visit http://www.uspto.gov/forms/. The filing date of the application will determine what form 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 http://www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

3. Claims 37-45, 55-63 and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-23 of U.S. Patent No. 10,287,631 (cited on IDS of 12/16/2019). Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims of U.S. Patent No. 10,287,631 represents a narrower species of the current claims and this species anticipates the current, more generic claims. Both sets of claims teach methods of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising the steps of amplifying each original strand of the double-stranded target nucleic acid molecule, sequencing the amplified target nucleic acid products generated from each strand, confirming the presence of at least one sequence read from each of the original strands, comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of both strands is complementary. The claims of U.S. Patent No. 10,287,631 also teach that the methods comprise ligating each of the double-stranded target nucleic acid molecules to at least one adapter molecule that comprises a degenerate or semi-degenerate single molecule identifier (SMI) sequence and a strand-distinguishing nucleotide sequence, grouping the first and second strand sequence

reads into a family of first and second strand sequence reads based on the SMI sequence alone or in combination with the target nucleic acid fragment ends, separating the first and second strand sequence reads into a set of first strand sequence reads and a set of second strand sequence reads based on the region of non-complementarity between the first and second strand of the adapter-target nucleic acid complex, and identifying nucleotide positions where the compared first and second strand sequence reads are non-complementary and where they are complementary.

4. Claims 37-45, 55-63 and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/118,290. Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims of copending U.S. Application No. 16/118,290 represents a narrower species of the current claims and this species anticipates the current, more generic claims. Both sets of claims teach methods of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising the steps of amplifying each original strand of the double-stranded target nucleic acid molecule, sequencing the amplified target nucleic acid products generated from each strand, confirming the presence of at least one sequence read from each of the original strands (see claim 41 of the '290 application), comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of both strands is

complementary (see claim 38 of the '290 application). The claims of copending U.S. Application No. 16/118,290 also teach that the methods comprise attaching adapters to both ends of double-stranded nucleic acid fragments in a sample to provide adapter-nucleic acid products, wherein the adapters comprise a double-stranded hybridized region, single-stranded 5' and 3' arms, and a physical unique molecular identifier (UMI) on at least one of the single-stranded arms that can be used to identify an individual molecule of a double-stranded nucleic acid fragment in the sample, sequencing the plurality of amplified polynucleotides to obtain a plurality of sequence reads each associated with a physical UMI, identifying a plurality of physical UMIs associated with the plurality of reads, and determining high accuracy sequence reads of the double-stranded nucleic acid fragments in the sample using the plurality of sequences obtained and the plurality of physical UMIs identified.

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

5. Claims 37-45, 55-63 and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/503,398. Although the conflicting claims are not identical, they are not patentably distinct from each other because the claims of copending U.S. Application No. 16/503,398 represents a narrower species of the current claims and this species anticipates the current, more generic claims. Both sets of claims teach methods of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising the steps of amplifying each original

strand of the double-stranded target nucleic acid molecule (see claim 40 of the '398 application), sequencing the amplified target nucleic acid products generated from each strand, confirming the presence of at least one sequence read from each of the original strands (see claim 55 of the '398 application), comparing the at least one sequence read from each strand with the at least one sequence read of the other strand, and generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position is identified as true if the particular position in the at least one sequence read of both strands is complementary. The claims of copending U.S. Application No. 16/503,398 also teach that the methods comprise preparing a sequencing library from a sample comprising a plurality of double-stranded DNA fragments by ligating adapter molecules to the plurality of double-stranded DNA fragments to generate a plurality of double-stranded adapter-DNA molecules).

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

Subject Matter Free of the Prior Art

6. With regard to claims 37-45, 55-63 and 65-67, no prior art was found that teaches or suggests a method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule, or a method of generating an error-corrected consensus sequence, as currently claimed. As discussed further below, the closest art of Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019) teaches methods for sequencing nucleic acid template constructs comprising double-stranded portions to provide redundant sequencing of both the sense

and antisense strands, wherein the constructs comprise a connecting hairpin adapter to enable sequencing of both the sense and antisense strands in the same sequencing template or contiguous nucleic acid molecule (see paragraphs 9, 54 and 139-141, and Figures 3A and 11). However, Travers does not teach a method that comprises "amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products" (independent claim 37) or generating a first set of duplicates of an original forward strand of the tagged DNA molecule and a second set of duplicates of an original complement strand of the tagged DNA molecule" (independent claim 55), as currently claimed. Rather, Travers teaches amplification and sequencing reactions that result in a continuous sequence of 700 bases representing both strands of a double-stranded molecule (see paragraph 141 and Figure 11) rather than generating distinct yet related sets of amplified molecules representing each strand. The secondary references of Xie et al. (U.S. Patent Pub. No. 2014/0155274, cited on IDS of 6/21/2019) and Bielas et al. (U.S. Patent Pub. No. 2015/0024950, cited on IDS of 6/21/2019) do not make up for the deficiencies of Travers.

Response to Arguments

7. Applicant's arguments filed February 18, 2021 have been fully considered and are persuasive.

Applicant requests that the rejections of claims 37-63, 65 and 66 under the judicially created doctrine of obviousness-type double patenting over claims 1-23 of U.S. Patent No. 10,287,631, claims 37-70 of copending Application No. 16/118,290 and

claims 37-70 of copending Application No. 16/503,398 be held in abeyance until at least one pending claim of the instant application is found to be otherwise allowable. Therefore, the rejections are maintained and have been updated to reflect the currently amended claims.

Applicant then argues that the objection of claims 55 and 59 for minor informalities should be withdrawn since the claims have been amended to correct the grammatical and spelling errors. The examiner agrees and therefore the objection is withdrawn.

Applicant then argues that the rejection of claim 58 under 35 U.S.C. § 112 (pre-AIA), second paragraph, as being indefinite should be withdrawn since the claims have been amended to remove the phrase "and preferably". The examiner agrees and therefore the rejection is withdrawn.

Applicant then argues that the rejection of claims 37-39, 41, 42, 45-58, 52-58 and 60-63 under 35 U.S.C. § 102(b) as being anticipated by Travers et al. (U.S. Patent Pub. No. 2009/0298075, cited on IDS of 6/21/2019) should be withdrawn since Travers fails to disclose or suggest each and every feature of independent claims 37 and 55. In particular, Applicant asserts that Travers does teach or suggest the limitation of "amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products" (claim 37) or generating a first set of duplicates of an original forward strand of the tagged DNA molecule and a second set of duplicates of an original complement strand of the tagged DNA molecule" (claim 55), as currently claimed, but rather teaches steps for performing single molecule sequencing reaction wherein sense

and antisense strands are linked together using hairpin adapters, thus resulting in a single set of amplified products from a double-stranded molecule. Applicant further argues that one of ordinary skill in the art would not be able to perform the additional steps of the claims if the amplifying or generating steps were carried out according to the teachings of Travers. The examiner agrees and therefore the rejections are withdrawn.

Applicant then argues that the rejections of claims 40 and 59 under 35 U.S.C. § 103(a) over Travers in view of Xie et al. (U.S. Patent Pub. No. 2014/0155274, cited on IDS of 6/21/2019), and claims 43, 44, 49-51, 65 and 66 over Travers in view of Bielas et al. (U.S. Patent Pub. No. 2015/0024950, cited on IDS of 6/21/2019), should be withdrawn since the primary reference of Travers fails to disclose all the elements recited in independent claims 37 and 55, and neither Xie nor Bielas make up for the deficiencies of Travers. The examiner agrees and therefore the rejections are withdrawn. It is further noted by Applicant that claims 49-51 have been canceled, and therefore the rejection of these claims under 35 U.S.C. § 103(a) is moot.

Summary

8. Claims 37-45, 55-63 and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting. However, claims 37-45, 55-63 and 65-67 are free of the prior art, as discussed above.

Conclusion

9. **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Correspondence

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to David C. Thomas whose telephone number is 571-272-3320 and whose fax number is 571-273-3320. The examiner can normally be reached on 5 days, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on 571-272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/David C Thomas/
Primary Examiner, Art Unit 1637

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

CPC - Searched*		
Symbol	**Date**	**Examiner**
C12Q 1/6806	11/10/2020	DCT
C12Q 1/6876	11/10/2020	DCT
C12Q 1/6869	11/10/2020	DCT
C12Q 2525/179	03/23/2021	DCT
C12Q 2525/191	03/23/2021	DCT
C12Q 2535/119	03/23/2021	DCT
C12Q 2535/122	03/23/2021	DCT
C12Q 2563/179	03/23/2021	DCT

CPC Combination Sets - Searched*		
Symbol	**Date**	**Examiner**

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	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

Search Notes		
Search Notes	**Date**	**Examiner**
EAST	11/09/2020	DCT
EAST	11/10/2020	DCT
EAST	11/12/2020	DCT
Google-search terms: sequencing double-stranded dna, both strands , barcodes, tags, y-shaped adapters, consensus sequences, sequence reads, single molecule identifier sequences, smi adapters, single strand consensus sequences	11/09/2020	DCT
PALM, inventor name search	11/09/2020	DCT
EAST-updated	03/23/2021	DCT
EAST-updated	03/24/2021	DCT

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

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. 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	16118306
Filing Date	2018-08-30
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	
Attorney Docket Number	072227-8043.US07

U.S.PATENTS Remove

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					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

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
/D.C.T/	1	20030165923	A1	2003-09-04	Li et al.	
/D.C.T/	2	20120094847	A1	2012-04-19	Warthmann et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	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]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16118306
Filing Date	2018-08-30
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	
Attorney Docket Number	072227-8043.US07

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, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
/D.C.T/	1	USPTO, Application No. 16/118,286, Non-Final Office Action, Examiner David C. THOMAS, mail date: 3/1/2021	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	03/23/2021

*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.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16118306
	Filing Date	2018-08-30
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	
	Attorney Docket Number	072227-8043.US07

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-03-22
Name/Print	Lara J. Dueppen	Registration Number	65002

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.**

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 the Freedom of Information Act requires disclosure of these record s.

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 inspections 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.

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L6	41,768	sequence read$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:48
L7	5,167	L6 and consensus sequence$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:48
L8	55	L7 and asymmetric adapt$3	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:49
L9	122	L7 and y-shaped adapt$3	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:50
L10	1	("20140155274").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/24 12:51
S1	1	("10287631").PN.	USPAT; USOCR	OR	OFF	2020/11/09 16:33
S2	1	("9752188").PN.	USPAT; USOCR	OR	OFF	2020/11/09 16:36
S3	40,249	sequence read$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:24
S4	0	S3 and amplif$7 each strand	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:25
S5	0	S3 and amplif$7 each original strand	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S6	19,195	S3 and amplif$7	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S7	4,281	S6 and consensus sequence	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S8	1,361	S7 and both strands	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:27
S9	1,283	S8 and double-stranded	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:28

S10	83	S9 and y-shaped adapt$2	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:28
S11	121	S8 and double-stranded sequencing	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:31
S12	513	c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:48
S13	183	c12q 1/6876	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S14	613	c12q 1/6869	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S15	84	S14 and c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S16	21	S13 and c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S17	27	S14 and c12q 1/6876	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:51
S18	1	("9970054").PN.	USPAT; USOCR	OR	OFF	2020/11/10 11:58
S19	1	("20190093161").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/10 11:59
S20	1	("20090298075").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/10 12:17
S21	1	("20150024950").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/12 18:21
S22	1	("20120058468").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 17:58
S23	1	("20140329282").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 18:00
S24	1	("20030165923").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 18:00
S25	1	("20120094847").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 18:01
S26	23	c12q 2525/179	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:08

S27	79	c12q 2525/191	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:08
S28	14	c12q 2535/119	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:09
S29	76	c12q 2535/122	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:09
S30	81	c12q 2563/179	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:09

3/24/2021 12:53:38 PM
C:\Users\dthomas1\Documents\EAST\Workspaces\16118306.wsp

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
/D.C.T/	A1	20120058468	A1	2012-03-08	McKeown	
/D.C.T/	A2	20140329282	A1	2015-01-22	Nelson et al.	

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]
/D.C.T/	C1	USPTO, Application No. 16/118,290, Non-Final Office Action, Examiner David C. THOMAS, mail date: 12/14/2020. 33 pages.	
/D.C.T/	C2	USPTO, Application No. 17/008,395, Non-Final Office Action, Examiner Stephanie Kane MUMMERT, mail date: 01/07/2021, 20 pages	
/D.C.T/	C3	USPTO, Application No. 16/118,305, Non-Final Office Action, Examiner David C. THOMAS, mail date: 1/13/2021	
/D.C.T/	C4	USPTO, Application No. 16/908,611, Notice of Allowance, Examiner David C. THOMAS, mail date: 1/29/2021. 11 pages.	
/D.C.T/	C5	USPTO, Application No. 16/119,471, Non-Final Office Action, Examiner David C THOMAS, mail date: 02/02/2021, 32 pages	

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	03/23/2021

*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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-02-18
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1	8318434	B2	2012-11-27	Cuppens	
	A2	9085798	B2	2015-07-21	Chee	
	A3	9783847	B2	2017-10-10	Chee	
	A4	10000800	B2	2018-06-19	Chee	
	A5	10119165	B2	2018-11-06	Chee	
	A6	10266883	B2	2019-04-23	Chee	
	A7	10266884	B2	2019-04-23	Chee	
	A8	10501793	B2	2019-12-10	Chee	

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
	A9	20050100900	A1	2005-05-12	Kawashima et al.	
	A10	20080261204	A1	2008-10-23	Lexow	
	A11	20160026758	A1	2016-01-28	Jabara et al.	
	A12	20200048701	A1	2020-02-13	Chee	
	A13	20200048702	A1	2020-02-13	Chee	
	A14	20200048703	A1	2020-02-13	Chee	

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]

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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]
	C1	USPTO, Application No. 17/008,395, Notice of Allowance, Examiner Stephanie Kane MUMMERT, mail date: 04/15/2021, 13 pages	
	C2	USPTO, Application No. 16/118,290, Notice of Allowance, Examiner David C. THOMAS, mail date: 4/20/2021. 5 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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-05-19
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
16/118,306	**05/19/2021 07:07:24 PM ET**	**072227-8043.US07**

Title of Invention

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	1114	FILED BY	Jonathan Petrovich
PATENT CENTER #	60198130	FILING DATE	08/30/2018
CUSTOMER #	94991	FIRST NAMED INVENTOR	Jesse Salk
CORRESPONDENCE ADDRESS	-	AUTHORIZED BY	Lara Dueppen

Documents

TOTAL DOCUMENTS: 11

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
2021-05-19 Supplemental IDS-197e1 072227-8043US07.pdf	3	Information Disclosure Statement (IDS) Form (SB08)	74 KB
Warning: This is not a USPTO supplied IDS fillable form. Data in the form cannot be automatically loaded to other USPTO systems.			
C01-Opposition-EP2828218-Withers-Rogers-2021-05-04.PDF	28	Non Patent Literature	3728 KB
C02-Opposition-EP2828218-Margaret-Dixon-2021-05-04.PDF	23	Non Patent Literature	3219 KB

C03-Opposition-EP2828218-Lewis-Silkin-2021-05-04.PDF	7	Non Patent Literature	996 KB
C04-Kinde-SI-2011-06-07.PDF	10	Non Patent Literature	1516 KB
C05-Schmitt-SI-2012-09-04.PDF	3	Non Patent Literature	497 KB
C06-Schmitt-Corrected-SI-2013-03-27.PDF	4	Non Patent Literature	562 KB
C07-Shiroguchi-SI-2012-01-24.PDF	8	Non Patent Literature	1425 KB
2021-05-19 Supplemental IDS-197e2 072227-8043US07.pdf	3	Information Disclosure Statement (IDS) Form (SB08)	73 KB

Warning: This is not a USPTO supplied IDS fillable form. Data in the form cannot be automatically loaded to other USPTO systems.

C1-US17008395-NOA-2021-04-15.PDF	13	Non Patent Literature	1327 KB
C2-US16118290_NOA_2019-04-20.PDF	5	Non Patent Literature	809 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
2021-05-19 Supplemental IDS-197e1 072227-8043US07.pdf	0DA65DC411B68DE058995783B23EAF2047154905BAE8F541B53B95E8938B7C8D87FE7A9764FDD03CE351579CEE07A2890E5521F0EFA5FDFA77D55151E8EE84F2
C01-Opposition-EP2828218-Withers-Rogers-2021-05-	1CB3811AB019C844FC215F9D4AC7731C42CA7E45A206DD25075D74E045F7B4D86F80C80DF417158AF21CB70D0952A349480

04.PDF	FD20FD36F37ABD963E0F164F1D585
C02-Opposition-EP2828218-Margaret-Dixon-2021-05-04.PDF	0B8069381E4DBB734A77DDD80B7592135CC2E3D465741257C20EF85EDE4ABA7DDA6257DA40E20EA441FE7B281C23AD481B5BB06B398C1E61B8ACB4A5170BFCE7
C03-Opposition-EP2828218-Lewis-Silkin-2021-05-04.PDF	3FA66133B4F0E2635F83895E8C8F4D33B248B63038921E95137C088BCA94A1127420E82F74018CB9C44C0052449D1E64580320A117FB8B67758FADC613B43EE9
C04-Kinde-SI-2011-06-07.PDF	8556FB268F7B4B9B19CA1130A53878971D7E6990720749C96F37B30EE9777E6BA66038EBABC14D0457B47212352F12587D036043DEFA4F32D2D5173403C089D1
C05-Schmitt-SI-2012-09-04.PDF	21C29E8444F64DB16B99D4AF5383BB2A84485E5A3F4D9F3EBF037E82E8FFADE0A027D068E49CF57243C6DB0A6BAE0E19F110CB3B1C8C97DAA8A308ADE30BF34D
C06-Schmitt-Corrected-SI-2013-03-27.PDF	E9C865F0ADD29F06D9CD666EE7DFAC0009CCB281FFE83BE1BEB5E3F4CAFAD868ED28DB0AA950CEB79E3AB8A14A300AEF69BEA8E8B909755F7171F172D64B7658
C07-Shiroguchi-SI-2012-01-24.PDF	1F560EA8259DD85D860430658ABC0E3F3E0CD359E262C9061202030B7F16829DCEB4200F88FF5A8BA9086001067A84BCBAA70DCC45B1EC23675359AD4E21B5F4
2021-05-19 Supplemental IDS-197e2 072227-8043US07.pdf	43DA64497E8532F406DD330BA046CAB2A91C5FEE6F0A77ECB42FDB78467F15D8130CB3DCF52569AD3EBB6EC72637EE83E3341336BA258913F1753C4E91DE0CD1
C1-US17008395-NOA-2021-04-15.PDF	172C8C0BF21895D739C259DF09B840EB3B2107DF91B7FBBA1344ECBD5840EB8A8E9AA2B5CD538AAD3F4CC0B0556B92CCD4524AF5672AC1B31675CEC77E18822A
C2-US16118290_NOA_2019-04-20.PDF	15F1F302A77319E8589F60129A273C2E0AC0376CA47DBCE411E1B63780FD129365E27EE247CC986191B3D0302C24EFC6567326A4453BD9FE0910318FCED2916D

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 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.

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1					

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
	A2					

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]
	C1	WITHERS & ROGERS, Opposition against European Patent No. 2828218, dated May 4, 2021, 28 pages.	
	C2	MARGARET DIXON, Opposition against European Patent No. 2828218, dated May 4, 2021, 23 pages.	
	C3	LEWIS SILKIN, Opposition against European Patent No. 2828218, dated May 4, 2021, 7 pages.	
	C4	Kinde et al., Supporting Information corresponding to "Detection and quantification of rare mutations with massively parallel sequencing", PNAS USA 108:9530-9535, 7 June 2011	
	C5	Schmitt et al., Supporting Information corresponding to "Detection of ultra-rare mutations by next-generation sequencing", PNAS, 4 Sep 2012, 109(36) 14508-14513	
	C6	Schmitt et al., Corrected Supporting Information corresponding to "Detection of ultra-rare mutations by next-generation sequencing", PNAS, 4 Sep 2012, 109(36) 14508-14513, Corrected SI dated 27 March 2013	
	C7	Shiroguchi et al., Supporting Information corresponding to "Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes", PNAS, ePub 109(4): 1347-1352), 9 Jan 2012	

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-05-19
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC PAYMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
16/118,306	05/19/2021 07:07:24 PM ET	072227-8043.US07

Title of Invention

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	1114	FILED BY	Jonathan Petrovich
PATENT CENTER #	60198130	AUTHORIZED BY	Lara Dueppen
CUSTOMER #	94991	FILING DATE	08/30/2018
CORRESPONDENCE ADDRESS	-	FIRST NAMED INVENTOR	Jesse Salk

Payment Information

PAYMENT METHOD	PAYMENT TRANSACTION ID	PAYMENT AUTHORIZED BY
CARD / 8735	E20215IJ07591709	Jonathan Petrovich

PRE-AUTHORIZED ACCOUNT	PRE-AUTHORIZED CATEGORY
500665	37 CFR 1.17 (Patent application and reexamination processing fees); 37 CFR 1.21 (Miscellaneous fees and charges)

FEE CODE	DESCRIPTION	ITEM PRICE($)	QUANTITY	ITEM TOTAL($)
1806	SUBMISSION OF AN INFORMATION DISCLOSURE STATEMENT	260.00	1	260.00
			TOTAL AMOUNT:	$260.00

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 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.

Doc Code: DIST.E.FILE Document Description: Electronic Terminal Disclaimer - Filed	PTO/SB/25 PTO/SB/26 U.S. Patent and Trademark Office Department of Commerce

Electronic Petition Request	**TERMINAL DISCLAIMER TO OBVIATE A PROVISIONAL DOUBLE PATENTING REJECTION OVER A PENDING "REFERENCE" APPLICATION AND TERMINAL DISCLAIMER TO OBVIATE A DOUBLE PATENTING REJECTION OVER A "PRIOR" PATENT**
Application Number	16118306
Filing Date	30-Aug-2018
First Named Inventor	Jesse Salk
Attorney Docket Number	072227-8043.US07
Title of Invention	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

☒ 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
UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION	100 %

The owner(s) of 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 any patent granted on pending reference Application Number(s)

16118290 filed on 08/30/2018

16503398 filed on 07/03/2019

as the term of any patent granted on said reference application may be shortened by any terminal disclaimer filed prior to the grant of any patent on the pending reference application. 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 any patent granted on the reference application 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 any patent granted on the instant application that would extend to the expiration date of the full statutory term of any patent granted on said reference application, "as the term of any patent granted on said reference application may be shortened by any terminal disclaimer filed prior to the grant of any patent on the pending reference application," in the event that any such patent granted on the pending reference application: 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 shortened by any terminal disclaimer filed prior to its grant.

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)

10287631

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.

Applicants 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 65002

( ) 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	/Lara J. Dueppen/

Name	Lara J. Dueppen

*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:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Filer:	Lara J. Dueppen/Amy Candeloro
Attorney Docket Number:	072227-8043.US07

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.: 16118306

Filing Date: 30-Aug-2018

Applicant/Patent under Reexamination: Salk

Electronic Terminal Disclaimer filed on June 29, 2021

☒ 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:	43124422
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Amy Candeloro
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	29-JUN-2021
Filing Date:	30-AUG-2018
Time Stamp:	20:44:44
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	E20216SK44350419
Deposit Account	500665
Authorized User	Amy Candeloro

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

37 CFR 1.17 (Patent application and reexamination processing 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	Terminal Disclaimer-Filed (Electronic)	eTerminal-Disclaimer.pdf	37418 35a6a6726ef559973b737cb844d567380b9081ce	no	3
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30577 44cfd973c2ea9b2fd510f963e262df0f8b613824	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	67995		

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.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	JESSE SALK, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1637
FILED:	AUGUST 30, 2018	EXAMINER:	DAVID C. THOMAS
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

RESPONSE TO FINAL OFFICE ACTION

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

Dear Commissioner:

In response to the Final Office Action mailed March 29, 2021, Applicant requests entry of the following:

A listing of the claims begins on page 2 of this paper.

Remarks begin on page 8 of this paper.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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
	A1					

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
	A2					

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]
	C1	USPTO, Application No. 16/119,471, Notice of Allowance, Examiner David C THOMAS, mail date: 05/24/2021, 11 pages	
	C2	USPTO, Application No. 16/503,398, Non-Final Office Action, Examiner Stephanie K. MUMMERT, mail date: 05/27/2021, 20 pages	
	C3	USPTO, Application No. 16/118,286, Notice of Allowance, Examiner David C. THOMAS, mail date: 06/23/2021, 8 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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-06-29
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

Electronic Patent Application Fee Transmittal

Application Number:	16118306
Filing Date:	30-Aug-2018
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Filer:	Lara J. Dueppen/Amy Candeloro
Attorney Docket Number:	072227-8043.US07

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:	43127083
Application Number:	16118306
International Application Number:	
Confirmation Number:	1114
Title of Invention:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING
First Named Inventor/Applicant Name:	Jesse Salk
Customer Number:	94991
Filer:	Lara J. Dueppen/Amy Candeloro
Filer Authorized By:	Lara J. Dueppen
Attorney Docket Number:	072227-8043.US07
Receipt Date:	29-JUN-2021
Filing Date:	30-AUG-2018
Time Stamp:	20:51:19
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	E20216SK52280475
Deposit Account	500665
Authorized User	Amy Candeloro

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

37 CFR 1.17 (Patent application and reexamination processing 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		2021-06-29-Response-072227-8043US07.pdf	157751 27351885b8b842d5a763fc2efb93490f33bea6ae	yes	10

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Applicant Arguments/Remarks Made in an Amendment	8	10
	Claims	2	7
	Response After Final Action	1	1

Warnings:

Information:

2	Information Disclosure Statement (IDS) Form (SB08)	2021-06-29-Supp-IDS-072227-8043US07.pdf	220351 b6d9a917064a8c44e2e0f89e4189488594d60fc4	no	2

Warnings:

Information:

This is not an USPTO supplied IDS fillable form

3	Other Reference-Patent/App/Search documents	2021-05-24-NOA-US16119471-8043US08.pdf	1506385 67a16b562604445e3306c8d6c47703dce8ab3aa5	no	11

Warnings:

Information:

4	Other Reference-Patent/App/Search documents	2021-05-27-NFOA-US16503398-8043US18.pdf	7343833 e8dda7c6e4a3f864ca537683fae866d726771653	no	42

Warnings:

Information:

5	Other Reference-Patent/App/Search documents	2021-06-23-NOA-US16118286-8043US04.pdf	1293538 1f949a19c043c55c93ebf9adf7aadddc3d5bf7bf	no	8
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	30842 71248160c1186b0d2e4bbf421f701fc2b90caba2	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	10552700		

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

Claims 37-45 and 55-67 were pending, with claim 64 withdrawn, when the present Office Action was mailed on March 29, 2021. Following this response, claims 37-45 and 55-67 will be pending and under examination.

Double Patenting

Claims 37-45, 55-63, and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-23 of U.S. Patent No. 10,287,631.

In the interest of expediting prosecution of the present application, and without agreeing or disagreeing with the rejection, Applicant submits a terminal disclaimer regarding U.S. Patent No. 10,287,631 herewith. In view of the foregoing, Applicant respectfully requests that the double patenting rejection of claims 37-45, 55-63, and 6-67 be withdrawn.

Claims 37-45, 55-63, and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/118,290.

In the interest of expediting prosecution of the present application, and without agreeing or disagreeing with the rejection, Applicant submits a terminal disclaimer regarding U.S. Application No. 16/118,290 herewith. In view of the foregoing, Applicant respectfully requests that the double patenting rejection of claims 37-45, 55-63, and 6-67 be withdrawn.

Claims 37-45, 55-63, and 65-67 are rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 37-70 of copending U.S. Application No. 16/503,398.

In the interest of expediting prosecution of the present application, and without agreeing or disagreeing with the rejection, Applicant submits a terminal disclaimer regarding U.S. Application No. 16/503,398 herewith. In view of the foregoing, Applicant respectfully requests that the double patenting rejection of claims 37-45, 55-63, and 6-67 be withdrawn.

The Applicant further notes that "the filing of a terminal disclaimer to obviate a rejection based on nonstatutory double patenting is not an admission of the propriety of the rejection." M.P.E.P. § 804.02(II) (citing *Quad Environmental Technologies Corp. v. Union Sanitary District*, 946 F.2d 870 (Fed. Cir. 1991) (The court indicated that the "filing of a terminal disclaimer simply serves the statutory function of removing the rejection of double patenting, and raises neither a presumption nor estoppel on the merits of the rejection.")).

Rejoinder of Withdrawn Claim 64

In view of the foregoing, Applicant submits that independent claim 55 is generic and is in condition for allowance. Thus, in accordance with MPEP § 821.04(a), Applicant respectfully requests rejoinder of withdrawn claim 64, which depends from or otherwise require all the limitations of allowable generic claim 55.

CONCLUSION

In view of the foregoing, Applicant requests reconsideration of the application and respectfully submits that the application is in condition for allowance. If the Examiner has any questions or believes a telephone conference would expedite prosecution of this application, the Examiner is encouraged to contact the undersigned at (310) 788-3349. Please charge any underpayment or credit any overpayment to our Deposit Account No. 50-0665.

Applicant reserves the right to later pursue any previously pending or other broader or narrow claims that capture any subject matter supported by the present disclosure, including subject matter that might be found disclaimed herein or by any prior prosecution. Accordingly, reviewers of this or any parent, child or related prosecution history shall not reasonably infer that the applicant has made any disclaimers or disavowals of any subject matter supported by the present application.

Dated: June 29, 2021

Respectfully submitted,

PERKINS COIE LLP

Customer No. 94991
UW/Perkins Coie LLP
Patent - LA
P.O. Box 1247
Seattle, WA 98111-1247
Phone: (310) 788-3349
Fax: (206) 332-7198

By: /Lara J. Dueppen/
Lara J. Dueppen
Registration No. 65,002

IN THE CLAIMS

The following is a complete listing of claims.

1-36. (Canceled)

37. (Previously Presented) A method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising:

amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products;

sequencing the amplified target nucleic acid products generated from each original strand;

confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands;

comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand; and

generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary.

38. (Previously Presented) The method of claim 37, wherein prior to amplifying each strand, the method further comprises ligating the double-stranded target nucleic acid molecule to an adapter molecule resulting in an asymmetric adapter-target nucleic acid complex such that each original strand generates a distinct yet related set of amplified target nucleic acid products during the amplification step.

39. (Previously Presented) The method of claim 38, wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

40. (Previously Presented) The method of claim 39, wherein the adapter molecule further comprises a nucleotide sequence that uniquely tags each original strand of the double-stranded target nucleic acid molecule such that each original strand is distinctly identifiable relative to its original complementary strand.

41. (Previously Presented) The method of claim 37, wherein the confirming step includes (i) grouping the sequence reads of the amplified target nucleic acid products into families of paired target nucleic acid strand sequences using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the double-stranded target nucleic acid molecule as an identifier, and (ii) separating each individual family of paired target nucleic acid strand sequences into a set of sequence reads from one original strand of the double-stranded target nucleic acid molecule and a set of sequence reads from the other original strand of the double stranded target nucleic acid molecule.

42. (Previously Presented) The method of claim 37, wherein the confirming step includes confirming the presence of at least about 3 sequence reads of an amplified target nucleic acid product generated from each of the original strands, and wherein the comparing step includes comparing a first consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from one original strand with a second consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from the other original strand.

43. (Previously Presented) The method of claim 37, further comprising identifying a mutation occurring at a particular position in the consensus sequence as a true mutation.

44. (Previously Presented) The method of claim 43, further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation.

45. (Previously Presented) The method of claim 37, wherein generating a consensus sequence of the double-stranded target nucleic acid molecule further comprises identifying one or more nucleotide positions that disagree between the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand, and discounting nucleotide positions in the consensus sequence that are in disagreement to provide an error-corrected sequence read of the double-stranded target nucleic acid molecule.

46-54. (Cancelled)

55. (Previously Presented) A method of generating an error-corrected consensus sequence comprising:

ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules;

for each of one or more tagged DNA molecules

generating a first set of duplicates of an original forward strand of the tagged DNA molecule and a second set of duplicates of an original complement strand of the tagged DNA molecule;

creating a single strand consensus sequence (SSCS) from the first set of duplicates of the original forward strand and a single strand

consensus sequence (SSCS) from the second set of duplicates of the original complement strand;

comparing the SSCS of the original forward strand to the SSCS of the original complement strand; and

generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward strand and the SSCS of the original complement strand are complementary.

56. (Previously Presented) The method of claim 55, wherein one or more adapter molecules are single molecule identifier (SMI) adapter molecules comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

57. (Previously Presented) The method of claim 56, wherein the SMI adapter molecules comprise a single-stranded SMI sequence or a double-stranded SMI sequence.

58. (Previously Presented) The method of claim 56, wherein the tagged DNA molecule comprises an SMI sequence in each of its original forward strand and original complement strand.

59. (Previously Presented) The method of claim 55, wherein the adapter molecules comprise a Y-shape, a U-shape, or a combination thereof.

60. (Previously Presented) The method of claim 55, wherein the tagged DNA molecule comprises an adapter molecule at each end.

61. (Previously Presented) The method of claim 55, wherein the comparing step includes pairing the SSCS of the original forward strand to the SSCS of the original

complement strand using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of a particular duplex DNA molecule as an identifier.

62. (Previously Presented) The method of claim 55, further comprising sequencing one or more of the duplicates of the original forward strand and one or more of the duplicates of the original complement strand of the tagged DNA molecules using a next-generation sequencing procedure.

63. (Previously Presented) The method of claim 55, wherein the first set of duplicates of the original forward strand and the second set of duplicates of the original complement strand includes a minimum of 1 member in each set.

64. (Withdrawn-Previously Presented) The method of claim 55, wherein the first set of duplicates of the original forward strand and the second set of duplicates of the original complement strand includes a minimum of 3 members in each set.

65. (Previously Presented) The method of claim 55, wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject, and wherein the method further comprises:

comparing the error-corrected consensus sequence to a respective reference sequence; and

detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

66. (Previously Presented) The method of claim 65, further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

67. (Previously Presented) The method of claim 58 wherein the SMI sequence in one strand comprises a first degenerate or semi-degenerate sequence and in the other strand comprises a second degenerate or semi-degenerate sequence that is complementary to the first such sequence.

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 16/118,306	Filing Date 08/30/2018	☐ 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 $460 =	
☐ 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	06/29/2021	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 22	Minus	** 30	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 2	Minus	*** 3	= 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".	/VICTOR BARLOW/
*** 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.

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

94991 7590 07/30/2021

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

EXAMINER
THOMAS, DAVID C

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 07/30/2021

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

TITLE OF INVENTION: METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

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	11/01/2021

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 and an extra copy of the form should be submitted. 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)

94991 7590 07/30/2021
Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

TITLE OF INVENTION: METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

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	11/01/2021

EXAMINER	ART UNIT	CLASS-SUBCLASS
THOMAS, DAVID C	1637	435-006120

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/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

94991 7590 07/30/2021

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 07/30/2021

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. 16/118,306	Applicant(s) Salk et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status No

-- 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 the submission of 29 June 2021.
   ☐ 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 37-45 and 55-67 . 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 6/29/2021, 5/19/2021, 5/19/2021.
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 _____.

/DAVID C THOMAS/
Primary Examiner, Art Unit 1637

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CPC							
Symbol						**Type**	**Version**
C12Q	/	1	/	6876		F	2013-01-01
C12Q	/	1	/	6806		I	2013-01-01
C12Q	/	1	/	6869		I	2013-01-01

CPC Combination Sets								
Symbol					**Type**	**Set**	**Ranking**	**Version**
C12Q	/	1	/	6869	I	1	1	2013-01-01
C12Q	/	2525	/	179	I	1	2	2013-01-01
C12Q	/	2525	/	185	I	1	3	2013-01-01
C12Q	/	2525	/	191	I	1	4	2013-01-01
C12Q	/	2535	/	119	I	1	5	2013-01-01
C12Q	/	1	/	6806	I	2	1	2013-01-01
C12Q	/	2525	/	191	I	2	2	2013-01-01
C12Q	/	2535	/	119	I	2	3	2013-01-01
C12Q	/	2535	/	122	I	2	4	2013-01-01
C12Q	/	2563	/	179	I	2	5	2013-01-01
C12Q	/	2565	/	514	I	2	6	2013-01-01

NONE (Assistant Examiner) (Date)	**Total Claims Allowed:** 22	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner) 23 July 2021 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure None

U.S. Patent and Trademark Office Part of Paper No.: 20210723

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q	/	1	/	6876
C12Q	/	1	/	6806
C12Q	/	1	/	6869

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**

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

NONE (Assistant Examiner) (Date)	**Total Claims Allowed:** 22	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner) 23 July 2021 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure None

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

☑ 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		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	22	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637	23 July 2021	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	None

U.S. Patent and Trademark Office Part of Paper No.: 20210723

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

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

CLAIM		DATE								
Final	Original	07/23/2021								
	1	-								
	2	-								
	3	-								
	4	-								
	5	-								
	6	-								
	7	-								
	8	-								
	9	-								
	10	-								
	11	-								
	12	-								
	13	-								
	14	-								
	15	-								
	16	-								
	17	-								
	18	-								
	19	-								
	20	-								
	21	-								
	22	-								
	23	-								
	24	-								
	25	-								
	26	-								
	27	-								
	28	-								
	29	-								
	30	-								
	31	-								
	32	-								
	33	-								
	34	-								
	35	-								
	36	-								
1	37	=								
2	38	=								
3	39	=								
4	40	=								
5	41	=								
6	42	=								

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CLAIM		DATE								
Final	Original	07/23/2021								
7	43	=								
8	44	=								
9	45	=								
	46	-								
	47	-								
	48	-								
	49	-								
	50	-								
	51	-								
	52	-								
	53	-								
	54	-								
10	55	=								
11	56	=								
12	57	=								
13	58	=								
14	59	=								
15	60	=								
16	61	=								
17	62	=								
18	63	=								
19	64	=								
20	65	=								
21	66	=								
22	67	=								

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

CPC - Searched*		
Symbol	**Date**	**Examiner**
C12Q 1/6806	11/10/2020	DCT
C12Q 1/6876	11/10/2020	DCT
C12Q 1/6869	11/10/2020	DCT
C12Q 2525/179	03/23/2021	DCT
C12Q 2525/191	03/23/2021	DCT
C12Q 2535/119	03/23/2021	DCT
C12Q 2535/122	03/23/2021	DCT
C12Q 2563/179	03/23/2021	DCT
C12Q 2525/185	07/22/2021	DCT
C12Q 2565/514	07/22/2021	DCT

CPC Combination Sets - Searched*		
Symbol	**Date**	**Examiner**

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	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

Search Notes		
Search Notes	**Date**	**Examiner**
EAST	11/09/2020	DCT
EAST	11/10/2020	DCT
EAST	11/12/2020	DCT
Google-search terms: sequencing double-stranded dna, both strands , barcodes, tags, y-shaped adapters, consensus sequences, sequence reads, single molecule identifier sequences, smi adapters, single strand consensus sequences	11/09/2020	DCT
PALM, inventor name search	11/09/2020	DCT
EAST-updated	03/23/2021	DCT
EAST-updated	03/24/2021	DCT
EAST-updated	07/22/2021	DCT

Interference Search			
US Class/CPC Symbol	**US Subclass/CPC Group**	**Date**	**Examiner**
	Claim search performed	07/23/2021	DCT

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	1	("20080261204").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/07/22 17:42
L2	1	("8318434").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/07/22 17:43
L3	1	("9085798").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/07/22 17:44
L4	1	("9783847").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/07/22 17:45
L5	1	("10266884").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/07/22 17:45
L6	1	("10501793").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/07/22 17:46
L8	31	c12q 2565/514	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/07/22 17:51
L9	10	c12q 2525/185	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/07/22 17:51
S1	1	("10287631").PN.	USPAT; USOCR	OR	OFF	2020/11/09 16:33
S2	1	("9752188").PN.	USPAT; USOCR	OR	OFF	2020/11/09 16:36
S3	40,249	sequence read$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:24
S4	0	S3 and amplif$7 each strand	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:25
S5	0	S3 and amplif$7 each original strand	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S6	19,195	S3 and amplif$7	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26
S7	4,281	S6 and consensus sequence	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:26

S8	1,361	S7 and both strands	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:27
S9	1,283	S8 and double-stranded	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:28
S10	83	S9 and y-shaped adapt$2	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:28
S11	121	S8 and double-stranded sequencing	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:31
S12	513	c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:48
S13	183	c12q 1/6876	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S14	613	c12q 1/6869	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S15	84	S14 and c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S16	21	S13 and c12q 1/6806	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:49
S17	27	S14 and c12q 1/6876	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2020/11/10 11:51
S18	1	("9970054").PN.	USPAT; USOCR	OR	OFF	2020/11/10 11:58
S19	1	("20190093161").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/10 11:59
S20	1	("20090298075").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/10 12:17
S21	1	("20150024950").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2020/11/12 18:21
S22	1	("20120058468").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 17:58
S23	1	("20140329282").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 18:00
S24	1	("20030165923").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 18:00

S25	1	("20120094847").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/23 18:01
S26	23	c12q 2525/179	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:08
S27	79	c12q 2525/191	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:08
S28	14	c12q 2535/119	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:09
S29	76	c12q 2535/122	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:09
S30	81	c12q 2563/179	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/23 18:09
S31	41,768	sequence read$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:48
S32	5,167	S31 and consensus sequence$1	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:48
S33	55	S32 and asymmetric adapt$3	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:49
S34	122	S32 and y-shaped adapt$3	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	NEAR	ON	2021/03/24 12:50
S35	1	("20140155274").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2021/03/24 12:51

7/22/2021 5:55:34 PM
C:\Users\dthomas1\Documents\EAST\Workspaces\16118306.wsp

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1					

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
	A2					

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]
	C1	WITHERS & ROGERS, Opposition against European Patent No. 2828218, dated May 4, 2021, 28 pages.	
	C2	MARGARET DIXON, Opposition against European Patent No. 2828218, dated May 4, 2021, 23 pages.	
	C3	LEWIS SILKIN, Opposition against European Patent No. 2828218, dated May 4, 2021, 7 pages.	
	C4	Kinde et al., Supporting Information corresponding to "Detection and quantification of rare mutations with massively parallel sequencing", PNAS USA 108:9530-9535, 7 June 2011	
	C5	Schmitt et al., Supporting Information corresponding to "Detection of ultra-rare mutations by next-generation sequencing", PNAS, 4 Sep 2012, 109(36) 14508-14513	
	C6	Schmitt et al., Corrected Supporting Information corresponding to "Detection of ultra-rare mutations by next-generation sequencing", PNAS, 4 Sep 2012, 109(36) 14508-14513, Corrected SI dated 27 March 2013	
	C7	Shiroguchi et al., Supporting Information corresponding to "Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes", PNAS, ePub 109(4): 1347-1352), 9 Jan 2012	

152539506.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

EXAMINER SIGNATURE			
Examiner Signature	/DAVID C THOMAS/	Date Considered	07/22/2021

*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. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-05-19
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

152539506.1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

0652

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
	A1	8318434	B2	2012-11-27	Cuppens	
	A2	9085798	B2	2015-07-21	Chee	
	A3	9783847	B2	2017-10-10	Chee	
	A4	10000800	B2	2018-06-19	Chee	
	A5	10119165	B2	2018-11-06	Chee	
	A6	10266883	B2	2019-04-23	Chee	
	A7	10266884	B2	2019-04-23	Chee	
	A8	10501793	B2	2019-12-10	Chee	

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
	A9	20050100900	A1	2005-05-12	Kawashima et al.	
	A10	20080261204	A1	2008-10-23	Lexow	
	A11	20160026758	A1	2016-01-28	Jabara et al.	
	A12	20200048701	A1	2020-02-13	Chee	
	A13	20200048702	A1	2020-02-13	Chee	
	A14	20200048703	A1	2020-02-13	Chee	

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]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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]
	C1	USPTO, Application No. 17/008,395, Notice of Allowance, Examiner Stephanie Kane MUMMERT, mail date: 04/15/2021, 13 pages	
	C2	USPTO, Application No. 16/118,290, Notice of Allowance, Examiner David C. THOMAS, mail date: 4/20/2021. 5 pages.	

EXAMINER SIGNATURE			
Examiner Signature	/DAVID C THOMAS/	Date Considered	07/22/2021

*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. /D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-05-19
Name/Print	Lara J. Dueppen	Registration Number	65,002

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. /D.C.T/

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	JESSE SALK, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1637
FILED:	AUGUST 30, 2018	EXAMINER:	DAVID C. THOMAS
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

RESPONSE TO FINAL OFFICE ACTION

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

Dear Commissioner:

In response to the Final Office Action mailed March 29, 2021, Applicant requests entry of the following:

A listing of the claims begins on page 2 of this paper.

Remarks begin on page 8 of this paper.

"OK to enter"

/D.C.T/

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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
	A1					

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
	A2					

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]
/D.C.T/	C1	USPTO, Application No. 16/119,471, Notice of Allowance, Examiner David C THOMAS, mail date: 05/24/2021, 11 pages	
/D.C.T/	C2	USPTO, Application No. 16/503,398, Non-Final Office Action, Examiner Stephanie K. MUMMERT, mail date: 05/27/2021, 20 pages	
/D.C.T/	C3	USPTO, Application No. 16/118,286, Notice of Allowance, Examiner David C. THOMAS, mail date: 06/23/2021, 8 pages	

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	07/22/2021

*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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-06-29
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 1114

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
16/118,306	08/30/2018	435	1637	072227-8043.US07

APPLICANTS

UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION, Seattle, WA

INVENTORS

Jesse Salk, Seattle, WA;
Lawrence A. Loeb, Bellevue, WA;
Michael Schmitt, Seattle, WA;

**** CONTINUING DATA ***************************

This application is a CON of 15/660,785 07/26/2017 PAT 10287631
which is a CON of 14/386,800 09/20/2014 PAT 9752188
which is a 371 of PCT/US2013/032665 03/15/2013
which claims benefit of 61/613,413 03/20/2012
and claims benefit of 61/625,623 04/17/2012
and claims benefit of 61/625,319 04/17/2012

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED ****

Foreign Priority claimed ☐ Yes ☑ No 35 USC 119(a-d) conditions met ☐ Yes ☐ No Verified and Acknowledged /DAVID C THOMAS/ Examiner's Signature	☐ Met after Allowance Initials	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
		WA	12	30	2

ADDRESS

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247
UNITED STATES

TITLE

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

FILING FEE RECEIVED		
1365	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	742	sequencing read$1.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:31
L2	11,788	double-stranded.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:32
L3	10,841	target nucleic acid$1.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:32
L4	1,701	high accuracy.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:32
L5	44	original strand.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:33
L6	381,267	amplif$7.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:33
L7	2,176	consensus sequence.clm.	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:34
L8	47	l1 same l2	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:34
L9	3	l3 same l4	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:34
L10	4	l3 same l5	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:34
L11	203	l6 same l7	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:35
L12	1	l8 same l11	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:35
L13	0	l8 same l9	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:37
L14	10	l8 same l3	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:38
L15	0	l5 same l14	US-PGPUB; USPAT	NEAR	ON	2021/07/23 17:38

7/23/2021 5:39:09 PM
C:\Users\dthomas1\Documents\EAST\Workspaces\16118306-2.wsp

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	JESSE SALK, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1637
FILED:	AUGUST 30, 2018	EXAMINER:	DAVID C. THOMAS
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

AMENDMENT UNDER 37 C.F.R. §1.312

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

Dear Commissioner:

In response to the Notice of Allowance mailed July 30, 2021, Applicant requests entry of the following:

A listing of the claims begins on page 2 of this paper.

Remarks begin on page 8 of this paper.

IN THE CLAIMS

The following is a complete listing of claims in the application.

1-36. (Canceled)

37. (Previously Presented) A method of generating a high accuracy sequence read of a double-stranded target nucleic acid molecule comprising:

amplifying each original strand of the double-stranded target nucleic acid molecule resulting in each original strand generating a distinct yet related set of amplified target nucleic acid products;

sequencing the amplified target nucleic acid products generated from each original strand;

confirming the presence of at least one sequence read of an amplified target nucleic acid product generated from each of the original strands;

comparing the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand; and

generating a consensus sequence of the double-stranded target nucleic acid molecule wherein each position in the consensus sequence is identified as true if the particular position in the at least one sequence read of both strands of the double-stranded target nucleic acid molecule is complementary.

38. (Previously Presented) The method of claim 37, wherein prior to amplifying each strand, the method further comprises ligating the double-stranded target nucleic acid molecule to an adapter molecule resulting in an asymmetric adapter-target nucleic acid complex such that each original strand generates a distinct yet related set of amplified target nucleic acid products during the amplification step.

39. (Previously Presented) The method of claim 38, wherein the adapter molecule is a single molecule identifier (SMI) adapter molecule comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

40. (Previously Presented) The method of claim 39, wherein the adapter molecule further comprises a nucleotide sequence that uniquely tags each original strand of the double-stranded target nucleic acid molecule such that each original strand is distinctly identifiable relative to its original complementary strand.

41. (Previously Presented) The method of claim 37, wherein the confirming step includes (i) grouping the sequence reads of the amplified target nucleic acid products into families of paired target nucleic acid strand sequences using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of the double-stranded target nucleic acid molecule as an identifier, and (ii) separating each individual family of paired target nucleic acid strand sequences into a set of sequence reads from one original strand of the double-stranded target nucleic acid molecule and a set of sequence reads from the other original strand of the double stranded target nucleic acid molecule.

42. (Previously Presented) The method of claim 37, wherein the confirming step includes confirming the presence of at least about 3 sequence reads of an amplified target nucleic acid product generated from each of the original strands, and wherein the comparing step includes comparing a first consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from one original strand with a second consensus sequence generated from the at least about 3 sequence reads of the amplified target nucleic acid product from the other original strand.

43. (Previously Presented) The method of claim 37, further comprising identifying a mutation occurring at a particular position in the consensus sequence as a true mutation.

44. (Previously Presented) The method of claim 43, further comprising identifying the double-stranded target nucleic acid molecule as being derived from a neoplastic cell by the true mutation.

45. (Previously Presented) The method of claim 37, wherein generating a consensus sequence of the double-stranded target nucleic acid molecule further comprises identifying one or more nucleotide positions that disagree between the at least one sequence read obtained from the amplified target nucleic acid products generated from one original strand with the at least one sequence read obtained from the amplified target nucleic acid products generated from the other original strand, and discounting nucleotide positions in the consensus sequence that are in disagreement to provide an error-corrected sequence read of the double-stranded target nucleic acid molecule.

46-54. (Cancelled)

55. (Previously Presented) A method of generating an error-corrected consensus sequence comprising:

ligating adapter molecules to individual duplex DNA molecules to form tagged DNA molecules;

for each of one or more tagged DNA molecules

generating a first set of duplicates of an original forward strand of the tagged DNA molecule and a second set of duplicates of an original complement strand of the tagged DNA molecule;

creating a single strand consensus sequence (SSCS) from the first set of duplicates of the original forward strand and a single strand

consensus sequence (SSCS) from the second set of duplicates of the original complement strand;

comparing the SSCS of the original forward strand to the SSCS of the original complement strand; and

generating an error-corrected consensus sequence having only nucleotide bases at which the sequence of both the SSCS of the original forward strand and the SSCS of the original complement strand are complementary.

56. (Previously Presented) The method of claim 55, wherein one or more adapter molecules are single molecule identifier (SMI) adapter molecules comprising an SMI sequence having at least one degenerate or semi-degenerate nucleic acid sequence.

57. (Previously Presented) The method of claim 56, wherein the SMI adapter molecules comprise a single-stranded SMI sequence or a double-stranded SMI sequence.

58. (Previously Presented) The method of claim 56, wherein the tagged DNA molecule comprises an SMI sequence in each of its original forward strand and original complement strand.

59. (Previously Presented) The method of claim 55, wherein the adapter molecules comprise a Y-shape, a U-shape, or a combination thereof.

60. (Previously Presented) The method of claim 55, wherein the tagged DNA molecule comprises an adapter molecule at each end.

61. (Previously Presented) The method of claim 55, wherein the comparing step includes pairing the SSCS of the original forward strand to the SSCS of the original

complement strand using a molecule-specific sequence at or adjacent to or at a specified position relative to an end of a particular duplex DNA molecule as an identifier.

62. (Previously Presented) The method of claim 55, further comprising sequencing one or more of the duplicates of the original forward strand and one or more of the duplicates of the original complement strand of the tagged DNA molecules using a next-generation sequencing procedure.

63. (Previously Presented) The method of claim 55, wherein the first set of duplicates of the original forward strand and the second set of duplicates of the original complement strand includes a minimum of 1 member in each set.

64. (Previously Presented) The method of claim 55, wherein the first set of duplicates of the original forward strand and the second set of duplicates of the original complement strand includes a minimum of 3 members in each set.

65. (Previously Presented) The method of claim 55, wherein at least some of the duplex DNA molecules are derived from a tumor or circulating neoplastic cells in a subject, and wherein the method further comprises:

comparing the error-corrected consensus sequence to a respective reference sequence; and

detecting the tumor or circulating neoplastic cells by identifying one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

66. (Previously Presented) The method of claim 65, further comprising identifying a drug-resistant variant of a cancer from the one or more nucleotide sequence variations present in the error-corrected consensus sequence and not present in the reference sequence.

67. (Currently Amended) The method of claim 58, wherein the SMI sequence in one strand comprises a first degenerate or semi-degenerate sequence and in the other strand comprises a second degenerate or semi-degenerate sequence that is complementary to the first such sequence.

REMARKS

Claims 37-45 and 55-67 were allowed prior to entry of the above-mentioned amendments. Claim 67 has been amended solely to add a comma to make the claim consistent with the other dependent claims. The foregoing amendments are made without prejudice to pursuing these claims in unamended or other forms in a continuation or other application. These amendments do not add any new matter or raise new issues. Accordingly, Applicant respectfully requests entry of these amendments under 37 C.F.R. §1.312.

Conclusion

If the Examiner has any questions, the Examiner is encouraged to contact the undersigned at (310) 788-3349. Please charge any underpayment or credit any overpayment to our Deposit Account No. 50-0665.

Although the present communication may include alterations to the application or claims, or characterizations of claim scope or referenced art, Applicant is not conceding that previously pending claims are not patentable.

Applicant reserves the right to later pursue any previously pending or other broader or narrow claims that capture any subject matter supported by the present disclosure, including subject matter that might be found disclaimed herein or by any prior prosecution. Accordingly, reviewers of this or any parent, child or related prosecution history shall not reasonably infer that the applicant has made any disclaimers or disavowals of any subject matter supported by the present application.

Dated: October 29, 2021

Respectfully submitted,

PERKINS COIE LLP

Customer No. 94991
UW/Perkins Coie LLP
Patent - LA
P.O. Box 1247
Seattle, WA 98111-1247
Phone: (310) 788-3349
Fax: (206) 332-7198

By: /Lara J. Dueppen/
Lara J. Dueppen
Registration No. 65,002

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
16/118,306	**10/29/2021 08:43:06 PM ET**	**072227-8043.US07**

Title of Invention

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	1114	FILED BY	Amy Candeloro
PATENT CENTER #	60377421	FILING DATE	08/30/2018
CUSTOMER #	94991	FIRST NAMED INVENTOR	Jesse Salk
CORRESPONDENCE ADDRESS	-	AUTHORIZED BY	Lara Dueppen

Documents

TOTAL DOCUMENTS: 6

DOCUMENT		PAGES	DESCRIPTION	SIZE (KB)
2021-10-29 312 Amendment 072227-8043.US07.pdf		8	-	100 KB
2021-10-29 312 Amendment 072227-8043.US07-A.NA.pdf	(1-1)	1	Amendment after Notice of Allowance (Rule 312)	85 KB
2021-10-29 312 Amendment 072227-8043.US07-CLM.pdf	(2-7)	6	Claims	90 KB
2021-10-29 312 Amendment 072227-8043.US07-REM.pdf	(8-8)	1	Applicant Arguments/Remarks Made in an Amendment	82 KB

2021-10-29 Issue Fee 072227-8043.US07.pdf	1	Issue Fee Payment (PTO-85B)	1608 KB
2021-10-29 Supplemental IDS 072227-8043.US07.pdf	2	Information Disclosure Statement (IDS) Form (SB08)	215 KB

Warning: This is not a USPTO supplied IDS fillable form. Data in the form cannot be automatically loaded to other USPTO systems.

US16118305 NOA 2021-09-23.pdf	12	Other Reference-Patent/App/Search documents	1582 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
2021-10-29 312 Amendment 072227-8043.US07.pdf	EFF849A021520F1C9B24A208B78091A139EF3ADBF2EF378817F72CC2228F0D3877B2D43D9EAB6F209C8E81F67AD9D83A85E117EED971E34DFB10B2E921743788
2021-10-29 312 Amendment 072227-8043.US07-A.NA.pdf	BD16CE32B6C3161CB6D6A4238D29758B622AD54B2EF0687E511198AA295238CABAC753E63E66A8D1EFB23C982DA59BC0DFFA56D12C392FF9885D146817A40A81
2021-10-29 312 Amendment 072227-8043.US07-CLM.pdf	AC9561842440DBDC9ECB68138780DF06496BBED45AAE0A8AB6D5F39F2E0F33AF768AE30C047E82392230577EBC68BBFA86118A05485D1B192C8573DCDB7EBAD8
2021-10-29 312 Amendment 072227-8043.US07-REM.pdf	FC9A64857B5FC0900E5AA7D759E5502CAD6A32051A5641ABB3F0852035E81AA222E3D014CED793033735C74FB6DEA8625E37325C6FE5759833F610EF557318FE
2021-10-29 Issue Fee 072227-8043.US07.pdf	812DD1CBD93274CA86680FEA20DA12E2FC2B79405D93E606CEB9A58CB160394F992EFD87A19F3C1F4CADBE50DF7B41D1B13A0A437E5DF7483C24B76EFAA698F6
2021-10-29 Supplemental IDS 072227-8043.US07.pdf	6CA8D22EF163F96A20E773208CC84103854BA9C344A1B36F5F4EE1F81967900F0EF9CC7E5E1448EAA52B518CD03E8E81D2

	94D0D5C968C67BB16762CF32D2A39A
US16118305 NOA 2021-09-23.pdf	055F301197096AB2AFC3DCE161CA92DB90E29C559A0A601AE9ED928FCAA0F187B19BD94A59A8D68647780705A68D87C107504D70A171114FD02F1CFCB9317E53

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 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.

UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC PAYMENT RECEIPT

APPLICATION #	RECEIPT DATE / TIME	ATTORNEY DOCKET #
16/118,306	10/29/2021 08:43:06 PM ET	072227-8043.US07

Title of Invention

METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

Application Information

APPLICATION TYPE	Utility - Nonprovisional Application under 35 USC 111(a)	PATENT #	-
CONFIRMATION #	1114	FILED BY	Amy Candeloro
PATENT CENTER #	60377421	AUTHORIZED BY	Lara Dueppen
CUSTOMER #	94991	FILING DATE	08/30/2018
CORRESPONDENCE ADDRESS	-	FIRST NAMED INVENTOR	Jesse Salk

Payment Information

PAYMENT METHOD	PAYMENT TRANSACTION ID	PAYMENT AUTHORIZED BY
DA / 500665	E20210SK48091153	Amy Candeloro

PRE-AUTHORIZED ACCOUNT	PRE-AUTHORIZED CATEGORY
500665	37 CFR 1.17 (Patent application and reexamination processing fees); 37 CFR 1.21 (Miscellaneous fees and charges)

FEE CODE	DESCRIPTION	ITEM PRICE($)	QUANTITY	ITEM TOTAL($)
1806	SUBMISSION OF AN INFORMATION DISCLOSURE STATEMENT	260.00	1	260.00
1501	UTILITY ISSUE FEE	1200.00	1	1200.00
			TOTAL AMOUNT:	$1,460.00

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 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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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
	A1					

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
	A2					

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]
	C1	USPTO, Application No. 16/118,305, Notice of Allowance, Examiner David C. THOMAS, mail date: 09/23/2021	

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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-10-29
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

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)

94991 7590 07/30/2021
Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

TITLE OF INVENTION: METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING

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	11/01/2021

EXAMINER	ART UNIT	CLASS-SUBCLASS
THOMAS, DAVID C	1637	435-006120

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/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-09 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 Perkins Coie LLP
2 Lara J. Dueppen
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: UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION

(B) RESIDENCE: (CITY and STATE OR COUNTRY): Seattle, WA

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. 500665

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 /Lara J. Dueppen/ Date October 29, 2021

Typed or printed name Lara J. Dueppen Registration No. 65,002

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.
16/118,306	08/30/2018	Jesse Salk	072227-8043.US07	1114

94991 7590 11/10/2021

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

EXAMINER	
THOMAS, DAVID C	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
11/10/2021	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):

patentprocurement@perkinscoie.com

Notice of Allowability	Application No. 16/118,306	Applicant(s) Salk et al.	
	Examiner DAVID C THOMAS	Art Unit 1637	AIA (FITF) Status No

-- 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 the submission of 29 October 2021.
   ☐ 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 37-45 and 55-67 . 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 10/29/2021.
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 _____.

/DAVID C THOMAS/
Primary Examiner, Art Unit 1637

Response to Rule 312 Communication	Application No.	Applicant(s)	
	16/118,306	Salk et al.	
	Examiner	Art Unit	AIA (FITF) Status
	DAVID C THOMAS	1637	No

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

1. ☑ The amendment filed on 29 October 2021 under 37 CFR 1.312 has been considered, and has been:

a) ☐ entered.

b) ☑ entered as directed to matters of form not affecting the scope of the invention.

c) ☐ disapproved because the amendment was filed after the payment of the issue fee.

Any amendment filed after the date the issue fee is paid must be accompanied by a petition under 37 CFR 1.313(c)(1) and the required fee to withdraw the application from issue.

d) ☐ disapproved. See explanation below.

e) ☐ entered in part. See explanation below.

/DAVID C THOMAS/
Primary Examiner, Art Unit 1637

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	Art Unit
	DAVID C THOMAS	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS

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

CLAIM		DATE								
Final	Original	07/23/2021	11/03/2021							
	1	-	-							
	2	-	-							
	3	-	-							
	4	-	-							
	5	-	-							
	6	-	-							
	7	-	-							
	8	-	-							
	9	-	-							
	10	-	-							
	11	-	-							
	12	-	-							
	13	-	-							
	14	-	-							
	15	-	-							
	16	-	-							
	17	-	-							
	18	-	-							
	19	-	-							
	20	-	-							
	21	-	-							
	22	-	-							
	23	-	-							
	24	-	-							
	25	-	-							
	26	-	-							
	27	-	-							
	28	-	-							
	29	-	-							
	30	-	-							
	31	-	-							
	32	-	-							
	33	-	-							
	34	-	-							
	35	-	-							
	36	-	-							
1	37	=	=							
2	38	=	=							
3	39	=	=							
4	40	=	=							
5	41	=	=							
6	42	=	=							

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CLAIM		DATE								
Final	Original	07/23/2021	11/03/2021							
7	43	=	=							
8	44	=	=							
9	45	=	=							
	46	-	-							
	47	-	-							
	48	-	-							
	49	-	-							
	50	-	-							
	51	-	-							
	52	-	-							
	53	-	-							
	54	-	-							
10	55	=	=							
11	56	=	=							
12	57	=	=							
13	58	=	=							
14	59	=	=							
15	60	=	=							
16	61	=	=							
17	62	=	=							
18	63	=	=							
19	64	=	=							
20	65	=	=							
21	66	=	=							
22	67	=	=							

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

CPC							
Symbol						**Type**	**Version**
C12Q	/	1	/	6876		F	2013-01-01
C12Q	/	1	/	6806		I	2013-01-01
C12Q	/	1	/	6869		I	2013-01-01

CPC Combination Sets								
Symbol						**Type**	**Set**	**Ranking**	**Version**
C12Q	/	1	/	6869		I	1	1	2013-01-01
C12Q	/	2525	/	179		I	1	2	2013-01-01
C12Q	/	2525	/	185		I	1	3	2013-01-01
C12Q	/	2525	/	191		I	1	4	2013-01-01
C12Q	/	2535	/	119		I	1	5	2013-01-01
C12Q	/	1	/	6806		I	2	1	2013-01-01
C12Q	/	2525	/	191		I	2	2	2013-01-01
C12Q	/	2535	/	119		I	2	3	2013-01-01
C12Q	/	2535	/	122		I	2	4	2013-01-01
C12Q	/	2563	/	179		I	2	5	2013-01-01
C12Q	/	2565	/	514		I	2	6	2013-01-01

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	22	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637	03 November 2021	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	None

U.S. Patent and Trademark Office — Part of Paper No.: 20211103

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

INTERNATIONAL CLASSIFICATION				
CLAIMED				
C12Q	/	1	/	6876
C12Q	/	1	/	6806
C12Q	/	1	/	6869

NON-CLAIMED				
	/		/	

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**

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

NONE (Assistant Examiner) (Date)		**Total Claims Allowed:** 22	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637 (Primary Examiner)	03 November 2021 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure None

U.S. Patent and Trademark Office

Part of Paper No.: 20211103

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/118,306	Salk et al.
	Examiner	**Art Unit**
	DAVID C THOMAS	1637

☑ 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		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	22	
/DAVID C THOMAS/ Primary Examiner, Art Unit 1637	03 November 2021	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	None

U.S. Patent and Trademark Office Part of Paper No.: 20211103

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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
	A1					

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
	A2					

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]
/D.C.T/	C1	USPTO, Application No. 16/118,305, Notice of Allowance, Examiner David C. THOMAS, mail date: 09/23/2021	

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	11/03/2021

*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.

Based on 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	16/118,306
	Filing Date	August 30, 2018
	First Named Inventor	Jesse Salk
	Art Unit	1637
	Examiner Name	David C. Thomas
	Attorney Docket Number	072227-8043.US07
	Confirmation Number	1114

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.

☒ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

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	/Lara J. Dueppen/	Date (YYYY-MM-DD)	2021-10-29
Name/Print	Lara J. Dueppen	Registration Number	65,002

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.**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF:	JESSE SALK, ET AL.	CONF. NO.:	1114
APPLICATION NO.:	16/118,306	ART UNIT:	1637
FILED:	AUGUST 30, 2018	EXAMINER:	DAVID C. THOMAS
FOR:	METHODS OF LOWERING THE ERROR RATE OF MASSIVELY PARALLEL DNA SEQUENCING USING DUPLEX CONSENSUS SEQUENCING		

AMENDMENT UNDER 37 C.F.R. §1.312

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

Dear Commissioner:

In response to the Notice of Allowance mailed July 30, 2021, Applicant requests entry of the following:

A listing of the claims begins on page 2 of this paper.

Remarks begin on page 8 of this paper.

"OK to enter"

/D.C.T/

Based on 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	16/118,306
Filing Date	August 30, 2018
First Named Inventor	Jesse Salk
Art Unit	1637
Examiner Name	David C. Thomas
Attorney Docket Number	072227-8043.US07
Confirmation Number	1114

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
/D.C.T/	A1	20120058468	A1	2012-03-08	McKeown	
/D.C.T/	A2	20140329282	A1	~~2015-01-22~~	Nelson et al. 11/2014	

Change(s) applied to document,
/J.M./
8/4/2021

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]
/D.C.T/	C1	USPTO, Application No. 16/118,290, Non-Final Office Action, Examiner David C. THOMAS, mail date: 12/14/2020. 33 pages.	
/D.C.T/	C2	USPTO, Application No. 17/008,395, Non-Final Office Action, Examiner Stephanie Kane MUMMERT, mail date: 01/07/2021, 20 pages	
/D.C.T/	C3	USPTO, Application No. 16/118,305, Non-Final Office Action, Examiner David C. THOMAS, mail date: 1/13/2021	
/D.C.T/	C4	USPTO, Application No. 16/908,611, Notice of Allowance, Examiner David C. THOMAS, mail date: 1/29/2021. 11 pages.	
/D.C.T/	C5	USPTO, Application No. 16/119,471, Non-Final Office Action, Examiner David C THOMAS, mail date: 02/02/2021, 32 pages	

EXAMINER SIGNATURE

Examiner Signature	/DAVID C THOMAS/	Date Considered	03/23/2021

*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.

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.
16/118,306	12/14/2021	11198907	072227-8043.US07	1114

94991 7590 11/23/2021

Perkins Coie LLP - University of Washington
PO Box 1247
Seattle, WA 98111-1247

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 337 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 Data Management (ODM) at (571)-272-4200.

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

UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION, Seattle, WA

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.